MAS223 Statistical Inference and Modelling

Exercises and Solutions

The exercises are grouped into sections, corresponding to chapters of the lecture notes. Within each section exercises are divided into warm-up questions, ordinary questions, and challenge questions. Note that there are no exercises accompanying Chapter 8.

The vast majority of exercises are ordinary questions. Ordinary questions will be used in homeworks and tutorials; they cover the material content of the course. Warm-up questions are typically easier, often nothing more than revision of relevant material from first year courses. Challenge questions are typically harder and test ingenuity.

This version of the exercises also contains solutions, which are written in blue. Solutions to challenge questions are not always included, hints may be given instead. Some of the solutions mention common pitfalls, written in red, which are mistakes that are (sometimes) easily made.

The solutions sometimes omit intermediate steps of basic calculations, which are left to the reader. For example, they may simply state $\int_0^x \lambda e^{-\lambda u} = 1 - e^{-\lambda u}$, and leave you to fill in the intermediate steps.

Contents

1 Univariate Distribution Theory

Warm-up Questions

1.1 Let X be a random variable taking values in $\{1,2,3\}$, with $\mathbb{P}[X=1]=\mathbb{P}[X=2]=0.4$. Find $\mathbb{P}[X=3]$, and calculate both $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

Solution. Since $\mathbb{P}[X=1]+\mathbb{P}[X=2]+\mathbb{P}[X=3]=1$, we have $\mathbb{P}[X=3]=0.2$. With this, we can calculate

$$\mathbb{E}[X]=1\mathbb{P}[X=1]+2\mathbb{P}[X=2]+3\mathbb{P}[X+3]=1.8$$
$$\mathbb{E}[X^2]=1^2\mathbb{P}[X=1]+2^2\mathbb{P}[X=2]+3^2\mathbb{P}[X+3]=3.8$$

Using that $\mathrm{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2$, we have $\mathrm{Var}(X)=0.56$.

1.2 Let Y be a random variable with probability density function (p.d.f.) $f(y)$ given by

$$f(y)=\begin{cases} y/2 & \text{for } 0\le y<2;\\ 0 & \text{otherwise.}\end{cases}$$

Find the probability that Y is between $\frac{1}{2}$ and 1. Calculate $\mathbb{E}[Y]$ and $\mathrm{Var}[Y]$.

Solution. We have $\mathbb{P}[Y\in[\frac{1}{2},1]]=\int_{1/2}^{1}(y/2)\,dy=3/16$. Similarly,

$$\mathbb{E}[Y]=\int_{-\infty}^{\infty}yf(y)=\int_0^2 y(y/2)\,dy=4/3$$
$$\mathbb{E}[Y^2]=\int_0^2(y^3/2)\,dy=2$$

so $\mathrm{Var}(Y)=\mathbb{E}[Y^2]-\mathbb{E}[Y]^2=2/9$.

Ordinary Questions

1.3 Define $F:\mathbb{R}\to[0,1]$ by

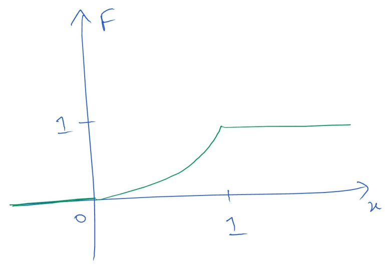
$$F(y)=\begin{cases} 0 & \text{for } y\le 0;\\ y^2 & \text{for } y\in(0,1);\\ 1 & \text{for } y\ge 1.\end{cases}$$

(a) Sketch the function F, and check that it is a distribution function.

(b) If Y is a random variable with distribution function F, calculate the p.d.f. of Y.

Solution.

(a) Sketch of F should look like



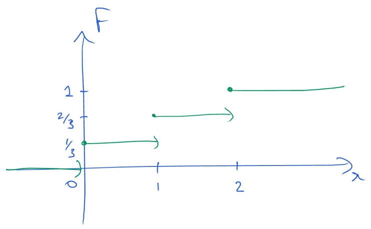
From the graph, F is continuous and (non-strictly) increasing. We have $F(x) = 0$ for all $x \le 0$, so $\lim_{x\to-\infty} F(x) = 0$. Similarly, $F(x) = 1$ for all $x \ge 1$, so $\lim_{x\to\infty} F(x) = 1$. Hence, F satisfies all the properties of a distribution function.

(b) We have $f(y) = F'(y)$, so treating each case in turn,

$$f(y) = \begin{cases} 0 & \text{for } y \le 0; \\ 2y & \text{for } y \in (0,1); \\ 0 & \text{for } y \ge 1. \end{cases}$$

1.4 Let X be a discrete random variable, taking values in $\{0,1,2\}$, where $\mathbb{P}[X = n] = \frac{1}{3}$ for $n \in \{0,1,2\}$. Sketch the distribution function $F_X : \mathbb{R} \to \mathbb{R}$.

Solution. Sketch should look like



Pitfall: The graph of F is not continuous; it jumps at $0,1,2$, and is otherwise constant.

1.5 Define $f : \mathbb{R} \to [0,1]$ by

$$f(x) = \begin{cases} 0 & \text{for } x < 0; \\ e^{-x} & \text{for } x \ge 0. \end{cases}$$

(a) Show that f is a probability density function.

(b) Find the corresponding distribution function and evaluate $\mathbb{P}[1 < X < 2]$.

Solution.

(a) Clearly $f(x) \ge 0$ for all x, and

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} e^{-x}\,dx = \left[-e^{-x}\right]_0^{\infty} = 1,$$

so f is a probability density function.

(b) We need to calculate $F(x) = \mathbb{P}[X \le x] = \int_{-\infty}^{x} f(u)\,du$. For $x \le 0$ we have $F(x) = \int_{-\infty}^{x} 0\,dx = 0$. For $x \ge 0$, we have $F(x) = \int_{-\infty}^{0} 0\,du + \int_{0}^{x} e^{-u}, du = 0 + (1 - e^{-x})$. Thus,

$$F(x) = \begin{cases} 0 & \text{for } x \le 0; \\ 1 - e^{-x} & \text{for } x \ge 0. \end{cases}$$

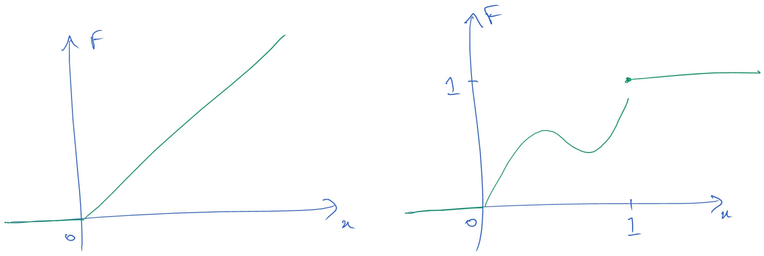
Hence, $\mathbb{P}[1 < X < 2] = \mathbb{P}[X < 2] - \mathbb{P}[X \le 1] = \mathbb{P}[X \le 2] - \mathbb{P}[X \le 1] = F(2) - F(1) = e^{-1} - e^{-2}$.

1.6 Sketch graphs of each of the following two functions, and explain why each of them is not a distribution function.

(a) $F(x) = \begin{cases} 0 & \text{for } x \le 0; \\ x & \text{for } x > 0. \end{cases}$

(b) $F(x) = \begin{cases} 0 & \text{for } x < 0; \\ x + \frac{1}{4}\sin 2\pi x & \text{for } 0 \le x < 1; \\ 1 & \text{for } x \ge 1. \end{cases}$

Solution. Sketches should look like



For (a), $F(x) > 1$ for $x > 1$, so F does not stay between 0 and 1. For (b), for $x \in [0, 1)$ we have $F'(x) = f(x) = 1 + \frac{2\pi}{4}\cos(2\pi x)$, which is negative at, for example, $x = \frac{1}{2}$, so (as is clear from the graph) F is not an increasing function.

1.7 Let $k \in \mathbb{R}$ and define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} k(x - x^2) & \text{for } x \in (0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Find the value of k for which $f(x)$ is a probability density function, and calculate the probability that X is greater than $\frac{1}{2}$.

Solution. We need $f(x) \ge 0$ for all x, so we need $k \ge 0$. Also, we need

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = k\int_0^1 x - x^2\,dx = k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{k}{6}.$$

So, $k = 6$. Therefore,

$$\mathbb{P}[X \ge \tfrac{1}{2}] = \int_{\frac{1}{2}}^{\infty} f(x)\,dx = \int_{\frac{1}{2}}^{1} 6(x - x^2)\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{\frac{1}{2}}^{1} = \frac{1}{2}.$$

1.8 The probability density function $f(x)$ is given by

$$f(x) = \begin{cases} 1+x & \text{for } -1 \le x \le 0; \\ 1-x & \text{for } 0 < x < 1; \\ 0 & \text{otherwise.} \end{cases}$$

Find the corresponding distribution function $F(x)$ for all real x.

Solution. We have $F(-1) = \mathbb{P}[X \le -1] = P[X < -1] = 0$. So $F(x) = 0$ for all $x \le -1$. If $x \in [-1, 0]$ then

$$F(x) = \int_{-\infty}^{x} f(u)\,du = F(-1) + \int_{-1}^{x} (1+u)\,du = 0 + \frac{(x+1)^2}{2}.$$

Now, for $x \in [0, 1]$, we have

$$F(x) = \int_{-\infty}^{x} f(u)\,du = F(0) + \int_{0}^{x} (1-u)\,du = \frac{1}{2} + x - \frac{x^2}{2} = \frac{1+2x-x^2}{2}.$$

Pitfall: Forgetting the F(0) results in missing out the term $\frac{1}{2}$. It needs to be present because for $x \in (0, 1)$ we have

$$F(x) = \mathbb{P}[X \le x] = \mathbb{P}[X \le 0] + \mathbb{P}[0 < X < x] = F(0) + \int_{0}^{x} (1-u)\,du.$$

Note that in the case of $x \in [-1, 0]$ the equivalent term was $F(-1)$ and was equal to 0.

Therefore, we have $F(1) = 1$. Since F is increasing and must stay between 0 and 1, we have $F(x) = 1$ for all $x \ge 1$.

Thus the distribution function $F(x)$ is

$$F(x) = \begin{cases} 0, & \text{for } x < -1 \\ \frac{(x+1)^2}{2}, & \text{for } -1 \le x < 0 \\ \frac{1+2x-x^2}{2}, & \text{for } 0 \le x < 1 \\ 1, & \text{for } x \ge 1 \end{cases}$$

1.9 Let

$$F(x) = \frac{e^x}{1+e^x} \quad \text{for all real } x.$$

(a) Show that F is a distribution function, and find the corresponding p.d.f. f.

(b) Show that $f(-x) = f(x)$.

(c) If X is a random variable with this distribution, evaluate $\mathbb{P}[|X| > 2]$.

Solution.

(a) Since $e^x \to 0$ as $x \to \infty$, we have

$$\lim_{x\to-\infty} \frac{e^x}{1+e^x} = \frac{0}{1+0} = 0$$
$$\lim_{x\to\infty} \frac{e^x}{1+e^x} = \lim_{x\to\infty} \frac{1}{e^{-x}+1} = \frac{1}{0+1} = 1$$

Pitfall: It does not make sense to use that $\lim_{x\to\infty} e^x = \infty$ and then incorrectly claim that $\frac{\infty}{1+\infty} = 1$.

Using the quotient rule, the derivative of f, the corresponding p.d.f. , is

$$f(x) = \frac{e^x(1+e^x) - e^x e^x}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2}.$$

Therefore, $f(x) > 0$ for all $x \in \mathbb{R}$, so F is an increasing function.
Since $x \mapsto e^x$ is continuous, F is a composition of sums and (non-zero) divisions of continuous functions; therefore F is continuous.
Hence, F satisfies all the properties of a distribution function.

(b) $f(-x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{2x}e^{-x}}{e^{2x}(1+e^{-x})^2} = \frac{e^x}{(e^x+1)^2} = f(x)$.

(c) We have

$$\begin{aligned}\mathbb{P}[|X| > 2] &= \mathbb{P}[X < -2] + \mathbb{P}[X > 2] \\ &= \mathbb{P}[X \le -2] + (1 - \mathbb{P}[X \le 2]) \\ &= 1 + F(-2) - F(2) \\ &= 1 + \tfrac{e^{-2}}{1+e^{-2}} - \tfrac{e^2}{1+e^2} \approx 0.238.\end{aligned}$$

Note that here we used $\mathbb{P}[X < -2] = \mathbb{P}[X \le -2]$, which holds because F is continuous.

1.10 Show that $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, defined for all $x \in \mathbb{R}$, is a probability density function.

Solution. Clearly $f(x) \ge 0$ for all x, and

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \frac{1}{\pi}\Big[\arctan(x)\Big]_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \frac{-\pi}{2}\right) = 1.$$

1.11 (a) For which values of $r \in [0,\infty)$ is $\int_1^\infty x^{-r}\,dx$ finite?

(b) Show that

$$f(x) = \begin{cases} x^{-2} & \text{if } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

is a probability density function.

(c) Show that the expectation of a random variable X, with the probability density function f given in (b), is not defined.

(d) Give an example of a random variable Y for which $\mathbb{E}[Y] < \infty$ but $\mathbb{E}[Y^2]$ is not defined.

Solution.

(a) For $r \neq -1$ we have

$$\int_1^\infty x^{-r} dx = \lim_{n\to\infty}\int_1^n x^{-r} dx = \lim_{n\to\infty} \frac{n^{-r+1}}{-r+1} - \frac{1}{-r+1}$$

which is finite (and equal to $\frac{1}{r-1}$) if $-r+1 < 0$ and infinite if $-r+1 > 0$. When $r = -1$ we have

$$\int_1^\infty x^{-1} dx = \lim_{n\to\infty}\int_1^n x^{-1} dx = \lim_{n\to\infty} \log n = \infty$$

Hence, $\int_1^\infty x^{-r} dx$ is finite if and only if $r > 1$.

(b) Clearly $f(x) \geq 0$ for all x and

$$\int_1^\infty x^{-2} dx = [(-x^{-1})]_{x=1}^\infty = 1$$

so f is a probability density function.

(c) we have $\int_{-\infty}^\infty x f_X(x) = \int_1^\infty x^{-1}\,dx$, which by part (a) is infinite. Hence the expectation of X is not defined.

(d) Let

$$f(y) = \begin{cases} 2y^{-3} & \text{if } x > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(y) \geq 0$ for all y, and

$$\int_{-\infty}^\infty f(y)\,dy = \int_1^\infty 2y^{-3}\,dy = \left[-y^{-2}\right]_{y=1}^\infty = 1$$

so f is a probability density function. From (a) we have that

$$\int_{-\infty}^\infty y^2 f_Y(y) dy = 2\int_1^\infty y^{-1} dy$$

is infinite and that

$$\int_{-\infty}^\infty y f_Y(y) dy = 2\int_1^\infty y^{-2} dy = \frac{2}{3}$$

is finite. Hence, if Y is a random variable with p.d.f. f, then $\mathbb{E}[Y]$ is defined but $\mathbb{E}[Y^2]$ is not.

1.12 The discrete random variable X has the probability function

$$\mathbb{P}[X = x] = \begin{cases} \frac{1}{x(x+1)} & \text{for } x \in \mathbb{N} \\ 0 & \text{otherwise.} \end{cases}$$

(a) Use the partial fractions of $\frac{1}{x(x+1)}$ to show that $\mathbb{P}[X \leq x] = 1 - \frac{1}{x+1}$, for all $x \in \mathbb{N}$.

(b) Write down the distribution function $F(x)$ of X, for $x \in \mathbb{R}$. Sketch its graph. What are the values of $F(2)$ and $F(\frac{3}{2})$?

(c) Evaluate $\mathbb{P}[10 \leq X \leq 20]$.

(d) Is $\mathbb{E}[X]$ defined? If so, what is $\mathbb{E}[X]$? If not, why not?

Solution.

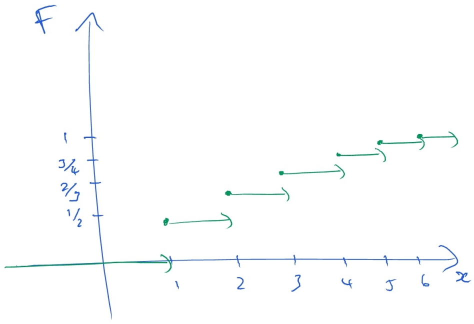
(a) Using partial fractions we obtain the identity $\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$, provided $x \neq 0, -1$. Hence, if $x \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}[X \leq x] &= \sum_{i=1}^{x} \mathbb{P}[X = x] \\
&= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \ldots + \left(\frac{1}{x} - \frac{1}{x+1}\right) \\
&= 1 - \frac{1}{x+1}.
\end{aligned}$$

(b) The distribution function is

$$F(u) = \begin{cases} 1 - \frac{1}{x+1}, & \text{for } u \in [x, x+1) \text{ where } x \in \mathbb{N} \\ 0, & \text{for } u < 1 \end{cases}$$

which looks like



$F(2) = 1 - \frac{1}{3} = \frac{2}{3}$ and $F(\frac{3}{2}) = F(1) = \frac{1}{2}$.

Pitfall: The graph of F is not continuous. The formula obtained in part (a) is only valid for $x \in \mathbb{N}$, and not for all $x \in \mathbb{R}$. Since X is a discrete random variable, its distribution function jumps at the points where X takes values (i.e. at $x \in \mathbb{N}$) and is constant in between those points.

(c) Since X is discrete,

$$\mathbb{P}[10 \le X \le 20] = \mathbb{P}[X \le 20] - \mathbb{P}[X \le 9] = F(20) - F(9) = \tfrac{11}{210}.$$

Pitfall: X is discrete, and $\mathbb{P}[X = 10] > 0$. So it's $F(20) - F(9)$, and not $F(20) - F(10)$.

(d) Since X is discrete, $\mathbb{E}[X]$ is defined if and only if $\sum_x x\mathbb{P}[X = x]$ converges. In our case, this sum is equal to

$$\sum_{x=1}^{\infty} x \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1}$$

which diverges.

To see that the sum diverges, we can write it as

$$\left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \dots$$

and note that each bracketed term is at least $\frac{1}{2}$; of course $\sum_{x=1}^{\infty} \frac{1}{2} = \infty$.

Challenge Questions

1.13 Show that there is no random variable X, with range $\mathbb{N}$, such that $\mathbb{P}[X = n]$ is constant for all $n \in \mathbb{N}$.

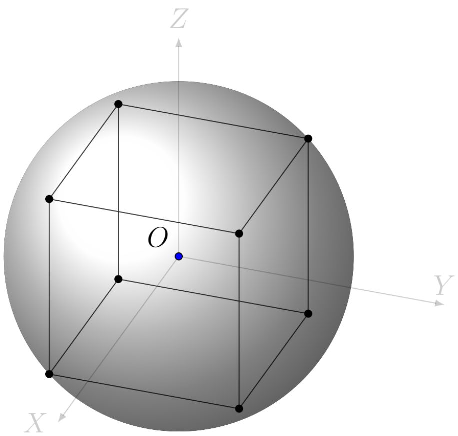
Solution. Since X has range $\mathbb{N}$ we have $\mathbb{P}[X \in \mathbb{N}] = 1$. If $\mathbb{P}[X = n] = \mathbb{P}[X = 1]$ for all n then we would have

$$1 = \mathbb{P}[X \in \mathbb{N}] = \sum_{n=1}^{\infty} \mathbb{P}[X = n] = \sum_{n=1}^{\infty} \mathbb{P}[X = 1].$$

If $\mathbb{P}[X = 1] = 0$ then $\mathbb{P}[X \in \mathbb{N}] = 0$, which is impossible, but similarly if $\mathbb{P}[X = 1] > 0$ then the sum is equal to $+\infty$, which is not possible either.

Hence, no such random variable exists.

1.14 Recall the meaning of 'inscribing' a cube within a sphere: the cube sits inside of the sphere, with each vertex of the cube positioned on the surface of the sphere. It is not especially easy to illustrate this on a two dimensional page, but here is an attempt:



Suppose that ten percent of the surface of the sphere is coloured blue, and the rest of the surface is coloured red. Show that, regardless of which parts are coloured blue, it is always possible to inscribe a cube within the sphere in such a way as all vertices of the cube are red.

Hint: A cube has eight corners. Suppose that position of the cube is sampled uniformly from the set of possible positions. What is the expected number of corners that are red?

Solution. Let X be a cube inscribed within the sphere, orientated uniformly at random. Strictly speaking, in three-dimensional spherical coordinates (r, θ, ϕ) this means we let the polar angle θ and azimuth angle ϕ be independent uniform random variables on $(0, 2\pi)$. More importantly, it means that the location of a given vertex of the cube is distributed uniformly on the surface of the sphere.

Let X be the number of corners that are red. Label the corners from $i = 1, \ldots, 8$. We can write

$$X = \sum_{i=1}^{8} \mathbb{1}_{\{A_i = \text{red}\}},$$

where A_i is the colour of the i^{th} corner. Here $\mathbb{1}_{\{A_i = \text{red}\}}$ is equal to 1 if A_i is red and equal to zero if A_i is blue. Hence,

$$\mathbb{E}[X] = \sum_{i=1}^{8} \mathbb{E}\left[\mathbb{1}_{\{A_i = \text{red}\}}\right] = \sum_{i=1}^{8} \mathbb{P}[A_i = \text{red}].$$

Since A_i is uniformly distributed on the surface of the sphere, and 90% of the sphere is red, we have $\mathbb{P}[A_i = \text{red}] = \frac{9}{10}$. Hence,

$$\mathbb{E}[X] = \tfrac{9}{10} \times 8 = 7.2.$$

Note that X can only take the values $\{1, 2, \ldots, 8\}$. Since $\mathbb{E}[X] > 7$, we must have $\mathbb{P}[X = 8] > 0$. Therefore, there are orientations of the cube for which all 8 vertices are red.

2 Standard Univariate Distributions

Warm-up Questions

2.1 (a) A standard fair dice is rolled 5 times. Let X be the number of sixes rolled. Which distribution (and which parameters) would you use to model X?

(b) A fair coin is flipped until the first head is shown. Let X be the total number of flips, including the final flip on which the first head appears. Which distribution (and which parameter) would you use to model X?

Solution.

(a) The binomial distribution, with parameters $n = 5$ and $p = \frac{1}{6}$.

(b) The geometric distribution, with parameter $p = \frac{1}{2}$.

Ordinary Questions

2.2 Let $\lambda > 0$. Write down the p.d.f. f of the random variable X, where $X \sim Exp(\lambda)$, and calculate its distribution function F. Hence, show that $\frac{f(t)}{1-F(t)}$ is constant for $t > 0$.

Solution. The p.d.f. of X is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0; \\ 0 & \text{otherwise.} \end{cases}$$

It's distribution function $F(x) = \int_{-\infty}^{x} f(u)\,du$ is clearly zero for $x \le 0$, and for $x > 0$ we have $\int_{-\infty}^{x} f(u)\,du = \int_0^x \lambda e^{-\lambda u}\,du = 1 - e^{-\lambda x}$. Therefore,

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x > 0; \\ 0 & \text{otherwise.} \end{cases}$$

Hence, for $t > 0$ we have $\frac{f(t)}{1-F(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$.

Pitfall: Don't forget the 'otherwise' case where $f(x) = 0$ or $F(x) = 0$. The same comment applies to many other questions.

2.3 Let $\lambda > 0$ and let X be a random variable with $Exp(\lambda)$ distribution. Let $Z = \lfloor X \rfloor$, that is let Z be X rounded down to the nearest integer. Show that Z is geometrically distributed with parameter $p = 1 - e^{-\lambda}$.

Solution. Since $X > 0$, we have $Z \in \{0, 1, 2, \ldots\}$, hence $\mathbb{P}[Z = z] = 0$ for all other z. For $n \in \{0, 1, 2, \ldots\}$ we have

$$\begin{aligned} \mathbb{P}[Z = n] &= \mathbb{P}[n \le X < n+1] \\ &= \int_n^{n+1} \lambda e^{-\lambda x}\,dx \\ &= e^{-\lambda n} - e^{-\lambda(n+1)} \\ &= e^{-\lambda n}(1 - e^{-\lambda}) \\ &= (1-p)^n p. \end{aligned}$$

which is the probability function of the geometric distribution.

2.4 Let $\mu \in \mathbb{R}$. Let X_1 and X_2 be independent random variables with distributions $N(\mu, 1)$ and $N(\mu, 4)$, respectively. Let T_1, T_2 and T_3 be defined by

$$T_1 = \frac{X_1 + X_2}{2}, \qquad T_2 = 2X_1 - X_2, \qquad T_3 = \frac{4X_1 + X_2}{5}.$$

Find the mean and variance of T_1, T_2 and T_3. Which of $\mathbb{E}[T_1]$, $\mathbb{E}[T_2]$ and $\mathbb{E}[T_3]$ would you prefer to use as an estimator of μ?

Solution. We have $\mathbb{E}[T_1] = \frac{1}{2}(\mathbb{E}[X_1] + \mathbb{E}[X_2]) = \mathbb{E}[T_2] = \mu$. Similarly, $\mathbb{E}[T_2] = \mathbb{E}[T_3] = \mu$, so all are unbiased when used as estimators of μ. We have

$$\mathrm{Var}(T_1) = (\tfrac{1}{2})^2 \left(\mathrm{Var}(X_1) + 2\,\mathrm{Cov}(X_1, X_2) + \mathrm{Var}(X_2)\right) = \tfrac{1}{4}(1 + 0 + 4) = \tfrac{5}{4},$$

and similarly $\mathrm{Var}(T_2) = 8$, $\mathrm{Var}(T_3) = \frac{4}{5}$.

On this information, we prefer $\mathbb{E}[T_3]$ as an estimator of μ, because T_3 has the smallest variance and so is likely to be closest to its mean.

2.5 Let X be a random variable with $Ga(\alpha, \beta)$ distribution.

(a) Let $k \in \mathbb{N}$. Show that

$$\mathbb{E}[X^k] = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}.$$

Hence, calculate $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$ and verify that these formulas match the ones given in lectures.

(b) Show that $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{2}{\sqrt{\alpha}}$.

Solution.

(a) From the p.d.f. of the Gamma distribution, we have

$$\begin{aligned}
\mathbb{E}[X^k] &= \int_0^\infty x^k \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{k+\alpha-1} e^{-\beta x}\, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(k+\alpha)}{\beta^{k+\alpha}} \\
&= \frac{\Gamma(k+\alpha)}{\beta^k \Gamma(\alpha)} \\
&= \frac{(k+\alpha-1)(k+\alpha-2)\cdots(k+\alpha-k+1)(k+\alpha-k)\Gamma(\alpha)}{\beta^k \Gamma(\alpha)} \\
&= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}.
\end{aligned}$$

To deduce the second line from the third line, we use Lemma 2.3 from lecture notes, and to deduce the fifth line from the fourth line we use Lemma 2.2, also from lecture notes.

Pitfall: It is true that $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$, but if $n \notin \mathbb{N}$ then $(n-1)!$ does not make sense. For general $\alpha \in (1, \infty)$ we have $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, which is what must be be used to deduce the fifth line above.

If $k = 1$ then $\mathbb{E}[X] = \frac{a}{b}$. If $k = 2$ then $\mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{\beta^2}$ and so the variance of X is $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$.

(b) With $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$, multiplying out and using the formulae from (a), we have

$$\begin{aligned}
\frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3} &= \frac{\mathbb{E}[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3]}{\sigma^3} \\
&= \frac{\mathbb{E}[X^3] - 3\mathbb{E}[X]\mathbb{E}[X^2] + 2\mathbb{E}[X]^3}{\sigma^3} \\
&= \frac{\frac{\alpha(\alpha+1)(\alpha+2)}{\beta^3} - \frac{3\alpha^2(\alpha+1)}{\beta^3} + \frac{2\alpha^3}{\beta^3}}{\sigma^3} \\
&= \frac{\alpha(\alpha^2 + 2\alpha + \alpha + 2 - 3\alpha^2 - 3\alpha + 2\alpha^2)/\beta^3}{\sqrt{\alpha^3}/\beta^3} \\
&= \frac{2}{\sqrt{\alpha}}.
\end{aligned}$$

2.6 (a) Using `R`, you can obtain a plot of, for example, the p.d.f. of a $Ga(3, 2)$ random variable between 0 and 10 with the command

```
curve(dgamma(x,shape=3,scale=2),from=0,to=10)
```

Use `R` to investigate how the shape of the p.d.f. of a Gamma distribution varies with the different parameter values. In particular, fix a value of β, see how the shape changes as you vary α.

(b) Investigate the effect that changing parameters values has on the shape of the p.d.f. of the Beta distribution. To produce, for example, a plot of the p.d.f. of $Be(4, 5)$, use

```
curve(dbeta(x,shape1=4,shape2=5),from=-1,to=2)
```

Solution.

(a) You should discover that decreasing α makes the p.d.f. appear more skewed (to the right). This makes it more likely that a sample of the random variable has a large value.

(b) You should discover that the parameter `shape1` (which we normally denote by α) controls the behaviour near $x = 0$, and `shape2` (that is, β), controls the behaviour near $x = 1$. In both case, the parameters can be tuned to cause (slow or fast) explosion to ∞, convergence to 1, and (slow or fast) convergence towards 0.

2.7 Suggest which standard discrete distributions (or combination of them) we should use to model the following situations.

(a) Organisms, independently, possess a given characteristic with probability p. A sample of k organisms with the characteristic is required. How many organisms will need to be tested to achieve this sample?

(b) In Texas Hold'em Poker, players make the best hand they can by combining two cards in their hand with five 'community' cards that are placed face up on the table. At the start of the game, a player can only see their own hand. The community cards are then turned over, one by one.

A player has two hearts in her hand. Three of the community cards have been turned over, and only one of them is a heart. How many hearts will appear in the remaining two community cards?

Use a computer to find the probability of seeing $k = 0, 1, 2$ hearts.

Solution.

(a) We'll need to sample, with success probability p, until we achieve k successes. So we will need to test $N \sim NegBin(k, p)$ organisms to find our sample.

(b) There are a total of 52 cards, 13 of each of the four suits. Our player can see 5 cards, 3 of which are hearts. Therefore, the unknown cards consist of 47 cards, 10 of which are hearts. The number of hearts that will be drawn in the next two community cards is, therefore, a hypergeometric distribution with parameters $N = 47$ (population size), $k = 10$ (successes), $n = 2$ (trials). As a result (use e.g. `R`), $\mathbb{P}[X = 0] \approx 0.65$, $\mathbb{P}[X = 1] \approx 0.32$ and $\mathbb{P}[X = 2] = 0.03$.

2.8 Let X be a $N(0, 1)$ random variable. Use integration by parts to show that $\mathbb{E}[X^{n+2}] = (n + 1)\mathbb{E}[X^n]$ for any $n = 0, 1, 2, \ldots$. Hence, show that

$$\mathbb{E}[X^n] = \begin{cases} 0 & \text{if } n \text{ is odd;} \\ (1)(3)(5)\ldots(n-1) & \text{if } n \text{ is even.} \end{cases}$$

Solution. Integrating by parts, for any $n \geq 0$, gives

$$\begin{aligned}
\mathbb{E}[X^n] &= \int_{-\infty}^{\infty} x^n \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\
&= \left[\frac{1}{\sqrt{2\pi}} \frac{x^{n+1}}{n+1} e^{-x^2/2}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{x^{n+1}}{n+1} \frac{1}{\sqrt{2\pi}} (-x) e^{-x^2/2}\, dx \\
&= 0 + \int_{-\infty}^{\infty} \frac{x^{n+2}}{n+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{n+1} \mathbb{E}[X^{n+2}]
\end{aligned}$$

Rearranging slightly,

$$\mathbb{E}[X^{n+2}] = (n+1)\mathbb{E}[X^n].$$

Since $\mathbb{E}[X] = 0$, induction gives that $\mathbb{E}[X^n] = 0$ for all odd n. Since $\mathbb{E}[X^0] = 1$, induction gives that $\mathbb{E}[X^n] = (1)(3)(5)\ldots(n-1)$ for all even n.

2.9 Let $X \sim N(\mu, \sigma^2)$. Show that $\mathbb{E}[e^X] = e^{\mu + \frac{\sigma^2}{2}}$.

Solution. Using the scaling properties of normal random variables from (2.2), we write $X = \mu + Y$ where $Y \sim N(0, \sigma^2)$. Hence, $e^X = e^{\mu+Y} = e^\mu e^Y$ and

$$\begin{aligned}
\mathbb{E}[e^Y] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^y e^{-\frac{y^2}{2\sigma^2}}\,dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left(y^2 + 2\sigma^2 y\right)\right\} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left((y+\sigma^2)^2 - \sigma^4\right)\right\} dy \\
&= e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{-\frac{(y+\sigma^2)^2}{2\sigma^2}\right\} dy \\
&= e^{\frac{\sigma^2}{2}}
\end{aligned}$$

Here, we use the same method as in Example 5: in the third line we complete the square and to deduce the final line we use that the p.d.f. of a $N(-\sigma^2, \sigma)$ random variable integrates to 1. Therefore,

$$\mathbb{E}[e^X] = e^\mu \mathbb{E}[e^Y] = e^{\mu + \frac{\sigma^2}{2}}.$$

Challenge Questions

2.10 Let X be a random variable with a continuous distribution, and a strictly increasing distribution function F. Show that $F(X)$ has a uniform distribution on $(0, 1)$.

Suggest how we might use this result to simulate samples from standard distributions.

Solution. Since F is strictly increasing, it has an inverse function F^{-1}. For $x \in (0, 1)$, we have

$$\mathbb{P}[F(X) \le x] = \mathbb{P}[X \le F^{-1}(x)] = F(F^{-1}(x)) = x.$$

Hence, $F(X)$ has the uniform distribution on $(0, 1)$.

We write this as $U = F(X)$, where U is uniform on $(0, 1)$. Therefore, $F^{-1}(U) = X$. Consequently, if we can simulate uniform random variables, and calculate $F^{-1}(x)$ for given x, we can simulate X as $F^{-1}(U)$.

In fact, this is a very common way of simulating random variables. Recall that a distribution function F is not necessarily strictly increasing, but it is necessarily *non-strictly* increasing. With some care, it is possible to extend this result to cover the general case. For many standard distributions, F^{-1} can be computed explicitly.

2.11 Prove that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Hint. Thanks to the normal distribution, you know that $\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$.

3 Transformations of Univariate Random Variables

Warm-up Questions

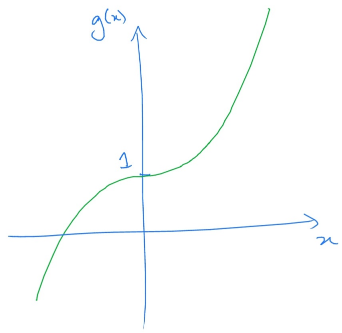
3.1 Let $g : \mathbb{R} \to \mathbb{R}$ be given by $g(x) = x^3 + 1$.

(a) Sketch the graph of g, show that g is strictly increasing, and find its inverse function g^{-1}.

(b) Let $R = [0, 2]$. Find $g(R)$.

Solution.

(a) A sketch of the function g looks like



It is clear from the graph that g is strictly increasing. If we set $y = x^3 + 1$ then $x = (y - 1)^{1/3}$, so the inverse function is $g^{-1}(y) = (y - 1)^{1/3}$.

(b) We have $g(0) = 1$ and $g(2) = 2^3 + 1 = 9$, so $g(R) = [1, 9]$.

Ordinary Questions

3.2 Let X be a random variable with p.d.f.

$$f_X(x) = \begin{cases} x^{-2} & \text{for } x > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Define $g(x) = e^x$ and let $Y = g(X)$.

(a) Show that $g(x)$ is strictly increasing. Find its inverse function $g^{-1}(y)$, and $\frac{dg^{-1}(y)}{dy}$.

(b) Identify the set R_X on which $f_X(x) > 0$. Sketch g and show that $g(R_X) = (e, \infty)$.

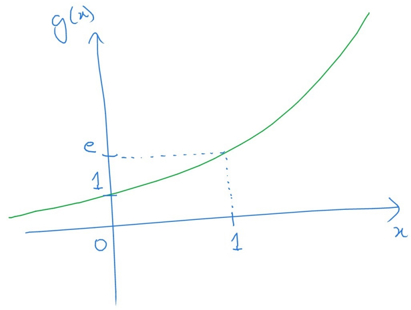
(c) Deduce from (a) and (b) that Y has p.d.f.

$$f_Y(y) = \begin{cases} (\log y)^{-2}\frac{1}{y} & \text{for } y > e \\ 0 & \text{otherwise.} \end{cases}$$

Solution.

(a) We have $\frac{dg(x)}{dx} = e^x > 0$, so g is strictly increasing. If we have $y = e^x$ then $x = \log y$, so the inverse function is $g^{-1}(y) = \log y$. Hence $\frac{dg^{-1}(y)}{dy} = \frac{1}{y}$.

(b) $f_X(x)$ is non-zero for $x \in R_X = (1, \infty)$. A sketch of g looks like



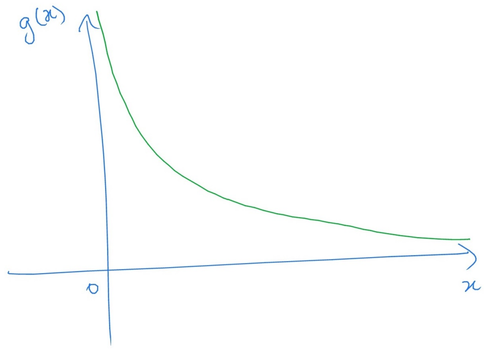
and hence $g(R_X) = (e, \infty)$.

(c) Since g is strictly increasing, we can use the formula

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \times \left|\frac{dg^{-1}(y)}{dy}\right| & \text{if } y \in g(R_X) \\ 0 & \text{otherwise,} \end{cases}$$
$$= \begin{cases} \frac{(\log y)^{-2}}{y} & \text{for } y > e \\ 0 & \text{otherwise.} \end{cases}$$

3.3 Let X be a random variable with the uniform distribution on $(0, 1)$, and let $Y = \frac{-\log X}{\lambda}$ where $\lambda > 0$. Show that Y has an $Exp(\lambda)$ distribution.

Solution. We have $f_X(x) = 1$ for $x \in (0, 1)$ (and $f_X(x) = 0$ otherwise). Hence $R_X = (0, 1)$. Our transformation is $g(x) = \frac{-\log x}{\lambda}$. For $x > 0$ we have $\frac{dg}{dx} = \frac{-1}{\lambda x} < 0$, so g is strictly decreasing. A sketch of g looks like



from which we can see that $g(R_X) = (0, \infty)$.

Writing $y = \frac{-\log x}{\lambda}$, we have $x = e^{-\lambda y}$ so $g^{-1}(y) = e^{-\lambda y}$ and $\frac{dg^{-1}}{dy} = -\lambda e^{-\lambda y}$. Hence, we can apply Lemma 3.1 to find $f_Y(y)$. Thus,

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left|\frac{dg^{-1}(y)}{dy}\right| & \text{for } y \in g(R_X) \\ 0 & \text{otherwise.} \end{cases}$$

Hence, using parts (a),(c) and (d) we obtain

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & \text{for } y \in (0,\infty); \\ 0 & \text{otherwise.} \end{cases}$$

This is the p.d.f. of the $Exp(\lambda)$ distribution. Hence, $Y \sim Exp(\lambda)$.

Pitfall: Don't forget to comment that g is strictly decreasing (or increasing), or else it isn't clear that you've checked whether or not Lemma 3.1 applies.

3.4 Let $\alpha, \beta > 0$.

(a) Show that $B(\alpha,\beta) = B(\beta,\alpha)$.

(b) Let X be a random variable with the $Be(\alpha,\beta)$ distribution. Show that $Y = 1 - X$ has the $Be(\beta,\alpha)$ distribution.

Solution.

(a) We have $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{\Gamma(\beta)\Gamma(\alpha)}{\Gamma(\alpha+\beta)} = B(\beta,\alpha)$.

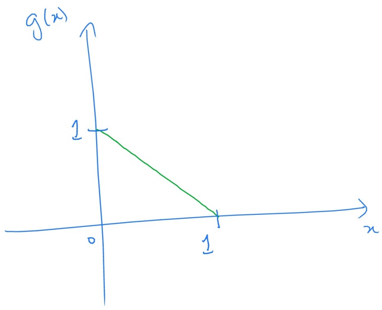
(b) We aim to use Lemma 3.1. We have

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } x \in (0,1); \\ 0 & \text{otherwise.} \end{cases}$$

Define $g : (0,1) \to (0,1)$ by

$$g(x) = 1 - x$$

and then $Y = g(X)$. Note that g is strictly increasing on $(0,1)$. We have $g^{-1}(y) = 1 - y$ and $\frac{dg^{-1}}{dy} = -1$. A sketch of g looks like



from which we can see that $g((0,1)) = (0,1)$. Hence, from Lemma 3.1 we have

$$f_Y(y) = f_X(g^{-1}(y)) \left|\frac{dg^{-1}}{dy}\right| = \frac{1}{B(\alpha,\beta)}(1-y)^{\alpha-1}y^{\beta-1} \times |-1|\,.$$

for $y \in (0,1)$, giving

$$f_Y(y) = \begin{cases} \frac{1}{B(\beta,\alpha)} y^{\beta-1}(1-y)^{\alpha-1} & \text{for } y \in (0,1); \\ 0 & \text{otherwise.} \end{cases}$$

This is the p.d.f. of the Beta(β,α) distribution.

3.5 Let $\alpha > 0$.

(a) Show that $B(\alpha, 1) = \frac{1}{\alpha}$.

(b) Let $X \sim Be(\alpha, 1)$ distribution. Let $Y = \sqrt[r]{X}$ for some positive integer r. Show that Y also has a Beta distribution, and find its parameters.

Solution.

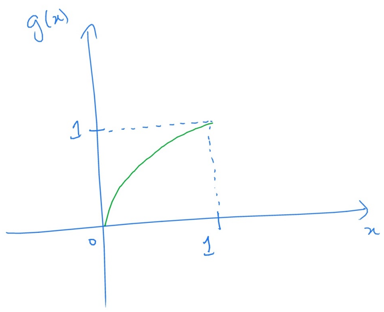
(a) From Lemma 2.2 we have $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, hence

$$B(\alpha, 1) = \frac{\Gamma(\alpha)\Gamma(1)}{\Gamma(\alpha+1)} = \frac{1}{\alpha}.$$

(b) We aim to use Lemma 3.1. From (a) we have,

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,1)} x^{\alpha-1} & \text{for } x \in (0,1), \\ 0 & \text{otherwise,} \end{cases}$$
$$= \begin{cases} \alpha x^{\alpha-1} & \text{for } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $R_X = (0, 1)$. The function $g(x) = \sqrt[r]{x}$



is strictly increasing on $(0, 1)$ and $g(R_X) = (0, 1)$. We have $g^{-1}(y) = y^r$ and $\frac{d}{dy} g^{-1}(y) = ry^{r-1}$. Hence, by Lemma 3.1, for $y \in (0, 1)$ we have

$$f_Y(y) = \alpha y^{r(\alpha-1)} \cdot ry^{r-1} = r\alpha y^{r\alpha-1}.$$

Thus,

$$f_Y(y) = \begin{cases} r\alpha y^{r\alpha-1} & \text{if } y \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

which is the p.d.f. of a $Be(r\alpha, 1)$ distribution. So Y has a Beta distribution with parameters $r\alpha$ and 1.

3.6 Let X have a uniform distribution on $[-1, 1]$. Find the p.d.f. of $|X|$ and identify the distribution of $|X|$.

Solution. If $x < 0$ then $\mathbb{P}[|X| \leq 0] = 0$. And since $|X| \in [0, 1]$, if $x > 1$ we have $\mathbb{P}[|X| \leq x] = 1$. For $x \in [0, 1]$ we have

$$\mathbb{P}[|X| \leq x] = \mathbb{P}[-x \leq X \leq x] = \int_{-x}^{x} f_X(u)\, du = \int_{-x}^{x} \frac{1}{2}\, du = x.$$

Differentiating, we have

$$f_{|X|}(x) = \begin{cases} 1 & \text{for } x \in [0,1]; \\ 0 & \text{otherwise.} \end{cases}$$

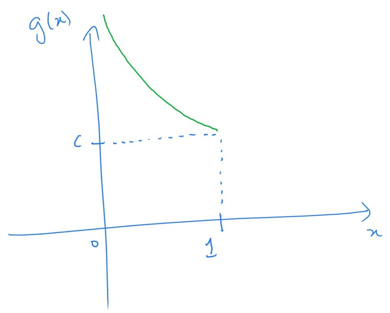
Therefore, $|X|$ is uniform on $[0,1]$.

Pitfall: If we try the 'standard' method of Lemma 3.1, we'll have $g(x) = |x|$, which is not monotone on $[-1,1]$; so the formula $f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}}{dy}|$ does not apply here – and it will give the wrong answer.

The same issue applies to several other questions in this section.

3.7 Let $\alpha, \beta > 0$ and let $X \sim Be(\alpha, \beta)$. Let $c > 0$ and set $Y = c/X$. Find the p.d.f. of Y.

Solution. We aim to use Lemma 3.1. We have $X \sim Be(\alpha, \beta)$ and $Y = c/X$, so set $g(x) = c/x$ where $g : (0,1) \to \mathbb{R}$.



Therefore, g is strictly decreasing on $(0,1)$. Then $g^{-1}(y) = \frac{c}{y}$, and $\frac{dg^{-1}}{dy} = \frac{-c}{y^2}$. Further, $f_X(x) > 0$ for $x \in (0,1)$, $R_X = (0,1)$ and $g(R_X) = (c, \infty)$. Hence, for $y > c$ we have

$$\begin{aligned} f_Y(y) &= \frac{1}{B(\alpha,\beta)} \left(\frac{c}{y}\right)^{\alpha-1} \left(1 - \frac{c}{y}\right)^{\beta-1} \frac{c}{y^2} \\ &= \frac{c^\alpha}{B(\alpha,\beta)} \frac{(y-c)^{\beta-1}}{y^{\alpha+\beta}}. \end{aligned}$$

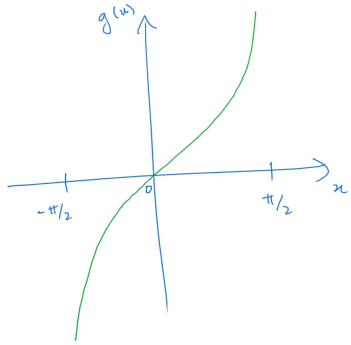
For $y \le c$, we have $f_Y(y) = 0$.

3.8 Let Θ be an angle chosen according to a uniform distribution on $(-\frac{\pi}{2}, \frac{\pi}{2})$, and let $X = \tan\Theta$. Show that X has the Cauchy distribution.

Solution. We aim to use Lemma 3.1 (with $X = \Theta$). The random variable Θ has a $U(-\frac{\pi}{2}, \frac{\pi}{2})$ distribution, so

$$f_\Theta(\theta) = \begin{cases} \frac{1}{\pi} & \text{for } -\frac{\pi}{2} < \theta < \frac{\pi}{2}; \\ 0 & \text{otherwise.} \end{cases}.$$

The function $g(\theta) = \tan\theta$ is strictly increasing on the range of Θ, that is on $(-\frac{\pi}{2}, \frac{\pi}{2})$.



Moreover, $g^{-1}(y) = \arctan(y)$ and $g^{-1}(y) \in (-\frac{\pi}{2}, \frac{\pi}{2})$ for all $y \in \mathbb{R}$. Hence, for all $y \in \mathbb{R}$,

$$f_Y(y) = f_\Theta(g^{-1}(y)) \left| \frac{dg^{-1}}{dy} \right| = \frac{1}{\pi} \frac{1}{1+y^2}$$

Hence, Y has the Cauchy distribution.

3.9 Let X be a random variable with the p.d.f.

$$f(x) = \begin{cases} 1+x & \text{for } -1 < x < 0; \\ 1-x & \text{for } 0 < x < 1; \\ 0 & \text{otherwise.} \end{cases}$$

Find the probability density functions of

(a) $Y = 5X + 3$

(b) $Z = |X|$

Solution.

(a) The function $g(x) = 5x + 3$ is strictly increasing, so we can use the formula

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

on each of the intervals in the definition of f_X. We note $g^{-1}(y) = \frac{y-3}{5}$ and $\frac{d}{dy} g^{-1}(y) = \frac{1}{5}$, so

$$\begin{aligned} f_Y(y) &= \begin{cases} \left(1 + \frac{y-3}{5}\right) \frac{1}{5} & \text{for } -1 < \frac{y-3}{5} < 0 \\ \left(1 - \frac{y-3}{5}\right) \frac{1}{5} & \text{for } 0 < \frac{y-3}{5} < 1 \\ 0 & \text{otherwise} \end{cases} . \\ &= \begin{cases} \frac{y+2}{25} & \text{for } -2 < y < 3 \\ \frac{8-y}{25} & \text{for } 3 < y < 8 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(b) The function $g(x) = |x|$ is not monotonic, so instead we use that for $z \geq 0$

$$\begin{aligned} F_Z(z) &= \mathbb{P}[Z \leq z] \\ &= \mathbb{P}[-z \leq X \leq z] \\ &= \begin{cases} \int_{-z}^{0}(1+u)\,du + \int_0^z (1-u)\,du & \text{for } z \leq 1 \\ 1 & \text{for } z > 1 \end{cases} \\ &= \begin{cases} 2z - z^2 & \text{for } z \leq 1 \\ 1 & \text{for } z > 1. \end{cases} \end{aligned}$$

If $z < 0$ then $\mathbb{P}[|X| < z] = 0$. So the p.d.f. of Z is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \begin{cases} 2(1-z) & 0 \leq z < 1 \\ 0 & \text{otherwise.} \end{cases}$$

3.10 Let X have the uniform distribution on $[a, b]$.

(a) For $[a, b] = [-1, 1]$, find the p.d.f. of $Y = X^2$.

(b) For $[a, b] = [-1, 2]$, find the p.d.f. of $Y = |X|$.

Solution.

(a) The probability density function of X is

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } x \in [-1, 1], \\ 0 & \text{otherwise.} \end{cases}$$

We have $Y \geq 0$, so $f_Y(y) = 0$ for $y \leq 0$. For $y \in (0, 1]$ we have

$$\mathbb{P}[Y \leq y] = \mathbb{P}[0 \leq X^2 \leq y] = \mathbb{P}[-\sqrt{y} \leq X \leq \sqrt{y}] = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\,dy = \sqrt{y}.$$

Thus, $\mathbb{P}[Y \leq 1] = 1$ and hence $\mathbb{P}[Y \leq y] = 1$ for all $y \geq 1$. Differentiating, we obtain

$$f_Y(y) = \begin{cases} 0 & \text{for } y \leq 0, \\ \frac{1}{2} y^{-1/2} & \text{for } y \in (0, 1], \\ 0 & \text{for } y > 1. \end{cases}$$

(b) The probability density function of X is

$$f_X(x) = \begin{cases} \frac{1}{3} & \text{for } x \in [-1, 2], \\ 0 & \text{otherwise.} \end{cases}$$

We have $Y \geq 0$, so $f_Y(y) = 0$ for $y \leq 0$. For $y > 0$, we need to consider three cases; $y \in (0, 1]$ and $y \in (1, 2]$ and $y > 2$.
For $y \in [0, 1]$, we have

$$\mathbb{P}[Y \leq y] = \mathbb{P}[-y \leq X \leq y] = \int_{-y}^{y} \frac{1}{3}\,dy = \frac{2y}{3}.$$

For $y \in (1,2]$, we have

$$\mathbb{P}[Y \leq y] = \mathbb{P}[Y \leq 1] + \mathbb{P}[1 < X \leq y] = \frac{2}{3} + \int_1^y \frac{1}{3}\, dy = \frac{2}{3} + \frac{y-1}{3}$$

For $y < 2$, we note that $\mathbb{P}[Y \leq 2] = 1$ from the previous case, so $\mathbb{P}[Y \leq y] = 1$ for all $y > 2$.

Differentiating, we obtain

$$f_Y(y) = \begin{cases} 0 & \text{for } y \leq 0 \text{ or } y > 2, \\ \frac{2}{3} & \text{for } y \in [0,1], \\ \frac{1}{3} & \text{for } y \in (1,2]. \end{cases}$$

3.11 Let X have a uniform distribution on $[-1,1]$ and define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \begin{cases} 0 & \text{for } x \leq 0; \\ x^2 & \text{for } x > 0. \end{cases}$$

Find the distribution function of $g(X)$.

Solution. Clearly $Y = g(X) \geq 0$, so $\mathbb{P}[Y \leq y] = 0$ for all $y < 0$. We have

$$\mathbb{P}[Y = 0] = \mathbb{P}[X \leq 0] = \int_{-1}^0 \frac{1}{2}\, dx = \frac{1}{2}.$$

For $y \in (0,1]$ we have

$$\mathbb{P}[Y \leq y] = \mathbb{P}[Y = 0] + \mathbb{P}[0 < Y \leq y] = \frac{1}{2} + \int_0^{\sqrt{y}} \frac{1}{2}\, dy = \frac{\sqrt{y}+1}{2},$$

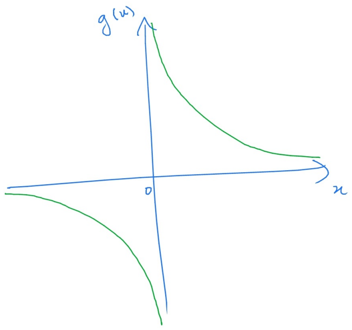
which means that $\mathbb{P}[Y \leq 1] = 1$ and hence $\mathbb{P}[Y \leq y] = 1$ for all $y \geq 1$.

To sum up,

$$F_Y(y) = \begin{cases} 0 & \text{for } y < 0, \\ \frac{1}{2} & \text{for } y = 0, \\ \frac{\sqrt{y}+1}{2} & \text{for } y \in (0,1], \\ 1 & \text{for } y > 1. \end{cases}$$

3.12 Let X be a random variable with the Cauchy distribution. Show that X^{-1} also has the Cauchy distribution.

Solution. The random variable X has p.d.f. $f_X(x) = \frac{1}{\pi(1+x^2)}$. Define $g : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$ by $g(x) = 1/x$.



Note that $\mathbb{P}[X=0]=0$, so the distribution of $Y=1/X$ is still well defined. We have $Y=g(X)$, but g is not strictly monotone (for example, $g(-1)<g(1)>g(2)$).

We plan to show that $\mathbb{P}[Y\leq y]=\mathbb{P}[X\leq y]$ for all $y\in\mathbb{R}$, which means that X and Y have the same distribution function; hence the same distribution.

Let us look at the case $y=0$ first. There is no value of $x\in\mathbb{R}$ such that $g(x)=0$, so $\mathbb{P}[Y=0]=0$. Since X is a continuous distribution, $\mathbb{P}[X=0]=0$. Also $X<0$ if and only if $Y<0$, so we have $\mathbb{P}[X\leq 0]=\mathbb{P}[X<0]=\mathbb{P}[Y<0]=\mathbb{P}[Y\leq 0]$.

For $y<0$, we have

$$\begin{aligned}\mathbb{P}[Y\leq y]&=\mathbb{P}\left[\tfrac{1}{y}\leq X<0\right]\\&=\int_{\frac{1}{y}}^{0}\frac{1}{\pi(1+x^2)}\,dx\\&=\int_{y}^{-\infty}\frac{1}{\pi(1+(1/v)^2)}\,\frac{-1}{v^2}\,dv\\&=\int_{-\infty}^{y}\frac{1}{\pi(1+v^2)}\,dv=\mathbb{P}[X\leq y].\end{aligned}$$

Here, we substitute $x=1/v$. For $y>0$, we must split into two cases, giving

$$\begin{aligned}\mathbb{P}[\tfrac{1}{X}\leq y]&=\mathbb{P}[X<0]+\mathbb{P}[\tfrac{1}{y}\leq X]\\&=\mathbb{P}[X<0]+\int_{\frac{1}{y}}^{\infty}\frac{1}{\pi(1+x^2)}\,dx\\&=\mathbb{P}[X<0]+\int_{y}^{0}\frac{1}{\pi(1+(1/v)^2)}\,\frac{-1}{v^2}\,dv\\&=\mathbb{P}[X<0]+\int_{0}^{y}\frac{1}{\pi(1+v^2)}\,dv\\&=\mathbb{P}[X<0]+\mathbb{P}[0\leq X\leq y]=\mathbb{P}[X\leq y].\end{aligned}$$

Again, we substitute $x=1/v$.

Hence, $\mathbb{P}[X\leq y]=\mathbb{P}[Y\leq y]$ for all $y\in\mathbb{R}$. Since X has a Cauchy distribution, so does Y.

Challenge Questions

3.13 If we were to pretend that $g(x)=1/x$ was strictly monotone, we could (incorrectly) apply Lemma 3.1 and use the formula $f_Y(y)=f_X(g^{-1}(y))|\frac{dg^{-1}}{dy}|$ to solve **3.12**. We would still arrive at the correct answer. Can you explain why?

Can you construct another example of a case in which the relationship $f_Y(y)=f_X(g^{-1}(y))|\frac{dg^{-1}}{dy}|$ holds, but where the function g is not monotone?

Hint. Carefully examine the proof of the formula for $f_Y(y)=f_X(g^{-1}(y))|\frac{dg^{-1}}{dy}|$ when *g is strictly monotone*, and compare it to the solution of **3.12**.

In general though, if g is not strictly monotone, the formula will not work!

3.14 Let Y and α,β be as in Question **3.7**.

(a) If $\alpha > 1$, show that $\mathbb{E}[Y] = \frac{c(\alpha+\beta-1)}{\alpha-1}$.

(b) If $\alpha \leq 1$ show that $\mathbb{E}[Y]$ is not defined.

Solution.

(a) We use that $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$. So

$$\begin{aligned}
\mathbb{E}\left[\frac{c}{X}\right] &= \int_{-\infty}^{\infty} \frac{c}{x} f_X(x)\, dx \\
&= \int_0^1 \frac{c}{x} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\, dx \\
&= \frac{c}{B(\alpha,\beta)} \int_0^1 x^{\alpha-2}(1-x)^{\beta-1}\, dx \\
&= \frac{cB(\alpha-1,\beta)}{B(\alpha,\beta)},
\end{aligned}$$

Expanding the Beta functions here in terms of the Gamma function, and using that $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, we get

$$c\frac{\frac{\Gamma(\alpha-1)\Gamma(\beta)}{\Gamma(\alpha+\beta-1)}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} = c\frac{\frac{\Gamma(\alpha-1)\Gamma(\beta)}{\Gamma(\alpha+\beta-1)}}{\frac{(\alpha-1)\Gamma(\alpha-1)\Gamma(\beta)}{(\alpha+\beta-1)\Gamma(\alpha+\beta-1)}} = \frac{c(\alpha+\beta-1)}{\alpha-1}.$$

(b) If we try to calculate $\mathbb{E}[Y]$ then, using the p.d.f. obtained in **3.7**(b) we want

$$\frac{c^\alpha}{B(\alpha,\beta)} \int_c^\infty \left(\frac{y-c}{y}\right)^{\beta-1} y^{-\alpha}\, dy. \tag{3.1}$$

We need to show that the integral does not converge. The idea is to note that $y - c \approx y$, for large y, which means that $(\frac{y-c}{y})^{\beta-1} \approx 1$; leaving us with $\int_c^\infty y^{-\alpha}\, dy$ which diverges to ∞.

To implement this idea, we could begin by noting that

$$(3.1) \geq \frac{c^\alpha}{B(\alpha,\beta)} \int_{c+1}^\infty \left(\frac{y-c}{y}\right)^{\beta-1} y^{-\alpha}\, dy \tag{3.2}$$

Since $y > c+1$, we have $\frac{1}{c+1} \leq \frac{y-c}{y} \leq 1$. If $\beta \geq 1$ then $(\frac{y-c}{y})^{\beta-1} \geq \frac{1}{(c+1)^{\beta-1}}$, and if $\beta \in (0,1]$ then $(\frac{y-c}{y})^{\beta-1} \geq 1$. Hence,

$$(3.2) \geq \frac{c^\alpha}{B(\alpha,\beta)} \min\left(\frac{1}{(c+1)^{\beta-1}}, 1\right) \int_{c+1}^\infty y^{-\alpha}\, dy.$$

Since $\int_{c+1}^\infty y^{-\alpha}\, dy$ diverges for $\alpha \leq 1$, we have that $\mathbb{E}[Y]$ does not exist for such α.

4 Multivariate Distribution Theory

Warm-up questions

4.1 Let $T = \{(x,y) \,:\, 0 < x < y\}$. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} e^{-2x-y} & \text{for } (x,y) \in T; \\ 0 & \text{otherwise.} \end{cases}$$

Sketch the region T. Calculate $\int_0^\infty \int_0^\infty f(x,y)\,dx\,dy$ and $\int_0^\infty \int_0^\infty f(x,y)\,dy\,dx$, and verify that they are equal.

Solution. See **4.2** for a sketch of T. We have

$$\begin{aligned} \int_{y=0}^\infty \int_{x=0}^\infty f(x,y)\,dx\,dy &= \int_{y=0}^\infty \int_{x=0}^y e^{-y}e^{-2x}\,dx\,dy = \int_0^\infty e^{-y}\left[\frac{-1}{2}e^{-2x}\right]_{x=0}^y dy \\ &= \int_0^\infty e^{-y}\left(\frac{-1}{2}e^{-2y} + \frac{1}{2}\right) dy = \frac{1}{2}\int_0^\infty e^{-y} - e^{-3y}\,dy \\ &= \frac{1}{2}\left[-e^{-y} + \frac{1}{3}e^{-3y}\right]_{y=0}^\infty = \frac{1}{2}\left(1 - \frac{1}{3}\right) = \frac{1}{3} \end{aligned}$$

and

$$\begin{aligned} \int_{x=0}^\infty \int_{y=0}^\infty f(x,y)\,dy\,dx &= \int_{x=0}^\infty \int_{y=x}^\infty e^{-y}e^{-2x}\,dx\,dy = \int_0^\infty e^{-2x}\left[-e^{-y}\right]_{y=x}^\infty dx \\ &= \int_0^\infty e^{-2x}e^{-x}\,dx = \int_0^\infty e^{-3x}\,dx = \left[\frac{-1}{3}e^{-3x}\right]_{x=0}^\infty = \frac{1}{3}, \end{aligned}$$

which are equal.

Pitfall: We must be careful to get the limits on the integrals correct. For $\int\int \dots dy\,dx$, we first allow x to vary between $0 \dots \infty$, which means that to cover T we must allow y to vary between x and ∞. We think of covering T by vertical lines, one for each $x = 0 \dots \infty$, each line with constant x and with y ranging from x up to ∞. It's helpful to draw a picture.

Alternatively, for $\int\int \dots dx\,dy$, we first allow y to vary between 0 and ∞, which means that to cover T we must allow x to vary between 0 and y. We think of covering T by horizontal lines, one for each $y = 0 \dots \infty$, each line with constant y and with x ranging from 0 up to y.

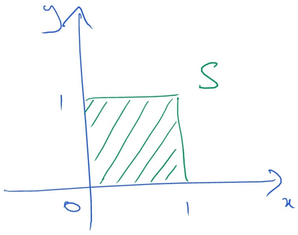
4.2 Sketch the following regions of $\mathbb{R}^2$.

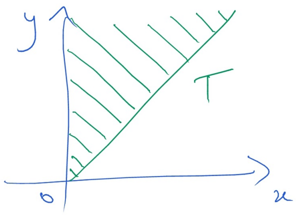
(a) $S = \{(x,y) \,:\, x \in [0,1],\, y \in [0,1]\}$.

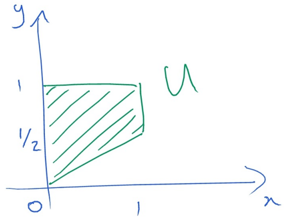
(b) $T = \{(x,y) \,:\, 0 < x < y\}$.

(c) $U = \{(x,y) \,:\, x \in [0,1],\, y \in [0,1],\, 2y > x\}$.

Solution.







Ordinary Questions

4.3 Let (X, Y) be a random vector with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} ke^{-(x+y)} & \text{if } 0 < y < x \\ 0 & \text{otherwise.} \end{cases}$$

(a) Using that $\mathbb{P}[(X, Y) \in \mathbb{R}^2] = 1$, find the value of k.

(b) For each of the regions S, T, U in **4.2**, calculate the probability that (X, Y) is inside the given region.

(c) Find the marginal p.d.f. of Y, and hence identify the distribution of Y.

Solution.

(a) Since $\mathbb{P}[(X, Y) \in \mathbb{R}^2] = 1$, we have $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\, dx = 1$. Therefore,

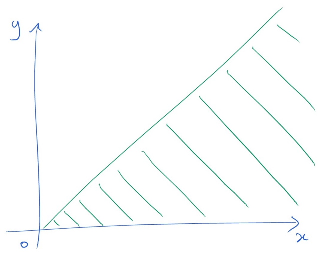
$$1 = k \int_0^{\infty} \int_0^{x} e^{-(x+y)}\, dy\, dx = k \int_0^{\infty} \left[-e^{-(x+y)} \right]_{y=0}^{x} dx = k \int_0^{\infty} (e^{-x} - e^{-2x})\, dx = \frac{k}{2}.$$

Hence, $k = 2$.

(We could also calculate $k \int_0^{\infty} \int_y^{\infty} e^{-(x+y)}\, dx\, dy$, with the same result.)

Pitfall: We must be careful to get limits of the inner integral correct. Allowing x to vary from $0 \ldots \infty$, and then allowing y to vary from $y = 0 \ldots x$, draws out precisely the range of (x, y) that make up $\{(x, y) \,:\, 0 < y < x\}$.

It's very helpful to draw a sketch of the region you're trying to integrate over:



See **4.1** for details of a similar case. The same issue applies to part (b) of this question, and to many others.

(b) We have

$$\mathbb{P}\left[(X,Y)\in S\right] = \int_0^1 \int_0^x 2e^{-(x+y)}\,dy\,dx = 2\int_0^1 e^{-x} - e^{-2x}\,dx = -2e^{-1} + e^{-2} + 1.$$

Since $f_{X,Y}(x,y) = 0$ for all $(x,y)\in T$, we have $\mathbb{P}[(X,Y)\in T] = 0$. Lastly,

$$\begin{aligned}\mathbb{P}\left[(X,Y)\in U\right] &= \int_0^1 \int_{x/2}^x 2e^{-(x+y)}\,dy\,dx \\ &= \int_0^1 2\left(-e^{-2x} + e^{-3x/2}\right)\,dx = e^{-2} - \frac{4}{3}e^{-3/2} + \frac{1}{3}.\end{aligned}$$

(In each case, we could instead calculate the $dx\,dy$ integral, with appropriate limits.)

(c) We integrate x out, giving, for $y > 0$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = 2\int_y^{\infty} e^{-(x+y)}\,dx = 2\left[-e^{-(x+y)}\right]_{x=y}^{\infty} = 2e^{-2y}.$$

So Y has an Exponential distribution with parameter $\lambda = 2$ (or, equivalently, a Gamma distribution with parameters $\alpha = 1$ and $\beta = 2$).

4.4 Let $S = [0,1]\times[0,1]$, and let U and V have joint probability density function

$$f_{U,V}(u,v) = \begin{cases} \frac{4u+2v}{3} & (u,v)\in S; \\ 0 & \text{otherwise.}\end{cases}$$

(a) Find $\mathbb{P}[U + V \le 1]$.

(b) Find $\mathbb{P}[V \le U^2]$.

Solution.

(a) We need to integrate the joint p.d.f. over the subset of (u,v) where $u + v \le 1$. Note that the joint p.d.f. is only non-zero when $u \ge 0$ and $v \ge 0$. So, we need to integrate over

$$\{(u,v)\ :\ 0 \le u \le 1,\ 0 \le v \le 1 - u\}$$

and we get

$$\begin{aligned}\mathbb{P}\left[U + V \le 1\right] &= \int_0^1 \int_0^{1-u} \frac{4u+2v}{3}\,dv\,du = \frac{1}{3}\int_0^1 (4u(1-u) + (1-u)^2)\,du \\ &= \frac{1}{3}\int_0^1 (1 + 2u - 3u^2)\,du = \frac{1}{3}.\end{aligned}$$

(b) We need to integrate the joint p.d.f. over the subset of (u,v) where $v \le u^2$. Note that the joint p.d.f. is only non-zero when $u\in[0,1]$ and $v\in[0,1]$. So, we need to integrate over

$$\{(u,v)\ :\ 0 \le u \le 1 \text{ and } 0 \le v \le u^2\}$$

and we get

$$\mathbb{P}\left[V \le U^2\right] = \int_0^1 \int_0^{u^2} \frac{4u+2v}{3}\,dv\,du = \frac{1}{3}\int_0^1 4u^3 + u^4\,du = \frac{2}{5}.$$

4.5 For the random variables U and V in Exercise **4.4**:

(a) Find the marginal p.d.f. $f_U(u)$ of U.

(b) Find the marginal p.d.f. $f_V(v)$ of V.

(c) For v such that $f_V(v) > 0$, find the conditional p.d.f. $f_{U|V=v}(u)$ of U given $V = v$.

(d) Check that each of f_U, f_V and $f_{U|V=v}$ integrate over $\mathbb{R}$ to 1.

(e) Calculate the two forms of conditional expectation, $\mathbb{E}[U|V = v]$, and $\mathbb{E}[U|V]$.

Solution.

(a) We must integrate v out of $f_{U,V}(u,v)$. For $u \in [0,1]$ this gives

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\,dv = \int_0^1 \frac{4u+2v}{3}\,dv = \frac{4u+1}{3}.$$

For $u \notin [0,1]$ we have $f_{U,V}(u,v) = 0$ so also $f_U(u) = 0$. Hence, the p.d.f. of U is

$$f_U(u) = \begin{cases} \frac{4u+1}{3} & \text{for } u \in [0,1]; \\ 0 & \text{otherwise.} \end{cases}$$

(b) Now, we integrate u out. For $v \in [0,1]$ we have

$$f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\,du = \int_0^1 \frac{4u+2v}{3}\,du = \frac{2v+2}{3}.$$

For $v \notin [0,1]$ we have $f_{U,V}(u,v) = 0$ so also $f_V(v) = 0$. Hence, the p.d.f. of V is

$$f_V(v) = \begin{cases} \frac{2v+2}{3} & \text{for } v \in [0,1]; \\ 0 & \text{otherwise.} \end{cases}$$

(c) When $f_V(v) > 0$, we have $f_{U|V=v}(u) = \frac{f_{U,V}(u,v)}{f_V(v)}$, which means that for $v \in [0,1]$ we have

$$f_{U|V=v}(u) = \begin{cases} \frac{2u+v}{v+1} & \text{for } u \in [0,1]; \\ 0 & \text{otherwise.} \end{cases}$$

(d) We check that $\int_0^1 \frac{4u+1}{3}\,du = \left[\frac{2u^2+u}{3}\right]_0^1 = 1$, and that $\int_0^1 \frac{2v+2}{3}\,dv = \left[\frac{2v^2+2v}{3}\right]_0^1 = 1$ and finally that $\int_0^1 \frac{2u+v}{1+v}\,du = \left[\frac{u^2+vu}{1+v}\right]_{u=0}^1 = 1$ as required.

(e) For $v \in [0,1]$ we have

$$\begin{aligned} g(v) = \mathbb{E}[U|V = v] &= \int_{-\infty}^{\infty} u f_{U|V=v}(u)\,du = \int_0^1 \frac{2u^2+uv}{1+v}\,du \\ &= \frac{1}{1+v}\left[\frac{2u^3}{3} + \frac{vu^2}{2}\right]_{u=0}^1 = \frac{1}{1+v}\left(\frac{2}{3} + \frac{v}{2}\right). \end{aligned}$$

Therefore, $\mathbb{E}[U|V] = g(V) = \frac{1}{1+V}\left(\frac{2}{3} + \frac{V}{2}\right)$.

4.6 Let (X, Y) be a random vector with joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{y-x}{2} & x \in [-1,0], y \in [0,1]; \\ \frac{x+y}{2} & x \in [0,1], y \in [0,1]; \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal p.d.f. of X. Show that the correlation coefficient X and Y is zero. Show also that X and Y are not independent.

Solution. The marginal p.d.f. of X is given by $f_X(x) = \int_0^1 f_{X,Y}(x,y)\, dy$. If $x \in [-1,0]$ then this is $\int_0^1 \frac{y-x}{2}\, dy = \frac{1-2x}{4}$, and if $x \in [0,1]$ it is $\int_0^1 \frac{x+y}{2}\, dy = \frac{1+2x}{4}$; otherwise it is zero. This gives $\mathbb{E}[X] = 0$, because $f_X(x) = f_X(-x)$.

To find $\mathbb{E}[XY]$ we calculate

$$\begin{aligned} \int_0^1 \int_{-1}^1 xy f_{X,Y}(x,y)\, dx\, dy &= \int_0^1 \int_{-1}^0 \frac{y^2x - x^2y}{2}\, dx\, dy + \int_0^1 \int_0^1 \frac{x^2y + y^2x}{2}\, dx\, dy \\ &= \int_0^1 \left(-\frac{y^2}{4} - \frac{y}{6} \right) dy + \int_0^1 \left(\frac{y^2}{4} + \frac{y}{6} \right) dy \\ &= 0. \end{aligned}$$

Hence, the covariance is $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$ (note that we do not need to know the value of $\mathbb{E}[Y]$). Hence the correlation coefficient is also zero.

To show dependence, we will show that the joint p.d.f. of X and Y does not factorise into the marginal densities. For example, dividing $f_{X,Y}(x,y)$ by $f_X(x)$ gives $\frac{2(y+x)}{1+2x}$ if $x \in [0,1]$ and $\frac{2(y-x)}{1-2x}$ if $x \in [-1,0]$, which does not depend only on y, so cannot be equal to $f_Y(y)$. Hence, X and Y are not independent.

4.7 Let X be a random variable. Let Z be a random variable, independent of X, such that $\mathbb{P}[Z = 1] = \mathbb{P}[Z = -1] = \frac{1}{2}$. Let $Y = XZ$.

(a) Show that X and Y are uncorrelated.

(b) Give an example in which X and Y are not independent.

Solution.

(a) We have

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X^2Z] - \mathbb{E}[X]\mathbb{E}[XZ] \\ &= \mathbb{E}[X^2]\mathbb{E}[Z] - \mathbb{E}[X]\mathbb{E}[X]\mathbb{E}[Z] \\ &= 0. \end{aligned}$$

Note that to deduce the second line we use the independence of X and Z, and to deduce the third line we use that $\mathbb{E}[Z] = (-1)\frac{1}{2} + (1)\frac{1}{2} = 0$.

(b) Let $\mathbb{P}[X = 0] = \mathbb{P}[X = 1] = \frac{1}{2}$, which means that $\mathbb{P}[Y = 1] = \mathbb{P}[Y = -1] = \frac{1}{2}$ and $\mathbb{P}[Y = 0] = \frac{1}{2}$. We have

$$\mathbb{P}[X = 1, Y = 1] = \mathbb{P}[X = 1, Z = 1] = \mathbb{P}[X = 1]\mathbb{P}[Z = 1] = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

$$\mathbb{P}[X = 1]\mathbb{P}[Y = 1] = \frac{1}{2}\frac{1}{4} = \frac{1}{8}.$$

Hence X and Y are not independent.

In fact, any example in which X is not equal to a constant will result in X and Y not being independent. Challenge question: prove this.

Pitfall: If X and Y are independent then $\mathrm{Cov}(X, Y) = 0$. However, there are many examples, such as this one, of cases in which $\mathrm{Cov}(X, Y) = 0$ where X and Y are not independent.

4.8 Let X and Y be independent random variables, with $0 < \mathrm{Var}(X) = \mathrm{Var}(Y) < \infty$. Let $U = X + Y$ and $V = XY$. Show that U and V are uncorrelated if and only if $\mathbb{E}[X] + \mathbb{E}[Y] = 0$.

Solution. Recall that U and V are uncorrelated if and only if $\mathrm{Cov}(U, V) = 0$. Using that X and Y are independent, we note that

$$\begin{aligned}
\mathrm{Cov}(U, V) &= \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V] \\
&= \mathbb{E}[(X + Y)XY] - \mathbb{E}[XY]\mathbb{E}[X + Y] \\
&= \mathbb{E}[X^2Y + Y^2X] - \mathbb{E}[X]\mathbb{E}[Y](\mathbb{E}[X] + \mathbb{E}[Y]) \\
&= \mathbb{E}[X^2]\mathbb{E}[Y] + \mathbb{E}[Y^2]\mathbb{E}[X] - \mathbb{E}[X]^2\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]^2 \\
&= \mathbb{E}[Y]\left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) + \mathbb{E}[X]\left(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2\right) \\
&= \mathbb{E}[Y]\,\mathrm{Var}(X) + \mathbb{E}[X]\,\mathrm{Var}(Y) \\
&= (\mathbb{E}[Y] + \mathbb{E}[X])\,\mathrm{Var}(X)
\end{aligned}$$

The last line follows because $\mathrm{Var}(X) = \mathrm{Var}(Y)$. Hence, U and V are uncorrelated if and only if $\mathbb{E}[X] = -\mathbb{E}[Y]$.

4.9 Let $\lambda > 0$. Let X have an $Exp(\lambda)$ distribution, and conditionally given $X = x$ let U have a uniform distribution on $[0, x]$. Calculate $\mathbb{E}[U]$ and $\mathrm{Var}(U)$.

Solution. Recall that the uniform distribution on $(0, x)$ has mean $\frac{x}{2}$ and variance $\frac{x^2}{12}$. We have $\mathbb{E}[U] = \mathbb{E}[\mathbb{E}[U|X]]$ and, since U is uniformly distributed on $(0, X)$ we have $\mathbb{E}[U|X] = X/2$. Hence, $\mathbb{E}[U] = \mathbb{E}[X/2] = \frac{1}{2\lambda}$.

Similarly, $\mathrm{Var}(U|X)$ is equal to the variance of a uniform distribution on $(0, X)$, which (from the sheet of distributions, or by a simple calculation) is $\frac{X^2}{12}$. Hence,

$$\mathrm{Var}(U) = \mathbb{E}[\mathrm{Var}(U|X)] + \mathrm{Var}(\mathbb{E}[U|X]) = \mathbb{E}\left[\frac{X^2}{12}\right] + \mathrm{Var}\left(\frac{X}{2}\right) = \frac{1}{12}\frac{2}{\lambda^2} + \frac{1}{4}\frac{1}{\lambda^2} = \frac{5}{12\lambda^2}$$

4.10 Let $k \in \mathbb{R}$ and let (X, Y) have joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} kx\sin(xy) & \text{for } x \in (0, 1), y \in (0, \pi), \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of k.

(b) For $x \in (0, 1)$, find the conditional probability density function of Y given $X = x$.

(c) Find $\mathbb{E}[Y|X]$.

Solution.

(a) We have $\mathbb{P}[(X,Y) \in \mathbb{R}^2] = 1$, so

$$\begin{aligned} 1 &= k \int_0^1 \int_0^\pi x \sin(xy)\, dy\, dx \\ &= k \int_0^1 [-\cos(xy)]_{y=0}^{\pi}\, dx \\ &= k \int_0^1 1 - \cos(\pi x)\, dx \\ &= k \left[x - \frac{1}{\pi} \sin(\pi x) \right]_{x=0}^1 = k. \end{aligned}$$

Hence, $k = 1$.

(b) First, we need the marginal density of X. For $y \in (0, \pi)$ we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_0^\pi x \sin(xy)\, dy = \frac{\pi}{2} [-\cos(xy)]_{y=0}^{\pi} = 1 - \cos(\pi x).$$

Hence,

$$f_X(x) = \begin{cases} 1 - \cos(\pi x) & \text{for } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, for $x \in (0,1)$, we have

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{x \sin(xy)}{1-\cos(\pi x)} & \text{for } y \in (0,\pi), \\ 0 & \text{otherwise.} \end{cases}$$

(c) Therefore, for $x \in (0,1)$ we have

$$\mathbb{E}[Y|X=x] = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)\, dy = \frac{x}{1-\cos(\pi x)} \int_0^\pi y \sin(xy)\, dy$$

To calculate the above, we use integration by parts to show that

$$\begin{aligned} \int_0^\pi y \sin(xy)\, dy &= \left[y \frac{-1}{x} \cos(xy) \right]_{y=0}^{\pi} + \int_0^\pi \frac{1}{x} \cos(xy)\, dy \\ &= \frac{\pi}{x} \cos(\pi x) + \left[\frac{1}{x^2} \sin(xy) \right]_{y=0}^{\pi} \\ &= \frac{\pi}{x} \cos(\pi x) + \frac{1}{x^2} \sin(\pi x). \end{aligned}$$

Hence,

$$\mathbb{E}[Y|X=x] = \frac{1}{1-\cos(\pi x)} \left(\pi \cos(\pi x) - \frac{1}{x} \sin(\pi x) \right)$$

which means that $\mathbb{E}[Y|X] = \frac{1}{1-\cos(\pi x)} \left(\pi \cos(\pi X) + \frac{1}{X} \sin(\pi X) \right)$.

4.11 Let U have a uniform distribution on $(0,1)$, and conditionally given $U = u$ let X have a uniform distribution on $(0,u)$.

(a) Find the joint p.d.f of (X,U) and the marginal p.d.f. of X.

(b) Show that $\mathbb{E}[U|X=x] = \frac{x-1}{\log x}$.

Solution.

(a) We have

$$f_{X|U}(x|u) = \begin{cases} 1/u & \text{for } x \in (0,u), \\ 0 & \text{otherwise,} \end{cases} \qquad f_U(u) = \begin{cases} 1 & \text{for } u \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

The joint pdf $f_{X,U}(x,u)$ satisfies $f_{X,U}(x|u) = \frac{f_{X|U}(x,u)}{f_U(u)}$ when $f_U(u) > 0$, hence

$$\begin{aligned} f_{X,U}(x,u) &= \begin{cases} 1/u & \text{for } u \in (0,1), x \in (0,u), \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} 1/u & \text{for } 0 < x < u < 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Hence, the marginal p.d.f. of X is

$$f_X(x) = \begin{cases} \int_x^1 1/u \, du & \text{for } x \in (0,1), \\ 0 & \text{otherwise,} \end{cases} \qquad = \begin{cases} -\log x & \text{for } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

(b) To calculate $\mathbb{E}[U|X=\frac{1}{2}]$, we need the conditional p.d.f. of U given $X = x$, which is

$$f_{U|X}(u|x) = \frac{f_{X,U}(x,u)}{f_X(x)} = \begin{cases} \frac{-1}{u \log x} & \text{for } u \in (x,1) \\ 0 & \text{otherwise,} \end{cases}$$

defined for $x \in (0,1)$. Hence,

$$\mathbb{E}\left[U|X=x\right] = \int_{-\infty}^{\infty} u f_{U|X}(u|x) = \int_x^1 u \frac{-1}{u \log(x)} \, du = \frac{x-1}{\log x}.$$

4.12 Let (X,Y) have a bivariate distribution with joint p.d.f. $f_{X,Y}(x,y)$. Let $y_0 \in \mathbb{R}$ be such that $f_Y(y_0) > 0$. Show that $f_{X|Y=y_0}(x)$ is a probability density function.

Solution. We must check that $f_{X|Y=y_0}$ is non-negative and integrates to 1.

Since $f_{X,Y}(x,y) \geq 0$ and $f_Y(y_0) > 0$, we have that $f_{X|Y=y_0}(x) = \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)} \geq 0$. Further,

$$\begin{aligned} \int_{-\infty}^{\infty} f_{X|Y=y_0}(x)\, dx &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)}\, dx \\ &= \frac{1}{f_Y(y_0)} \int_{-\infty}^{\infty} f_{X,Y}(x,y_0)\, dx \\ &= \frac{1}{f_Y(y_0)} f_Y(y_0) = 1, \end{aligned}$$

as required. Here we use the definition of the marginal distribution $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$.

Challenge Questions

4.13 Give an example of random variables (X, Y, Z) such that

$$\mathbb{P}[X < Y] = \mathbb{P}[Y < Z] = \mathbb{P}[Z < X] = \frac{2}{3}.$$

Solution. Let X be uniformly distributed on the three element set $\{0, 1, 2\}$. Set $Y = (X + 1) \mod 3$, and $Z = (Y + 1) \mod 3$. Of course, then $X = (Z + 1) \mod 3$.

The real puzzle, of course, is how to dream up this example (or another that does the same job).

This question is related to the observation that, in an election between three candidates A, B, C, it is possible for more than half of the voters to prefer A to B, for more than half to prefer B to C, and for more than half to prefer C to A.

5 Transformations of Multivariate Distributions

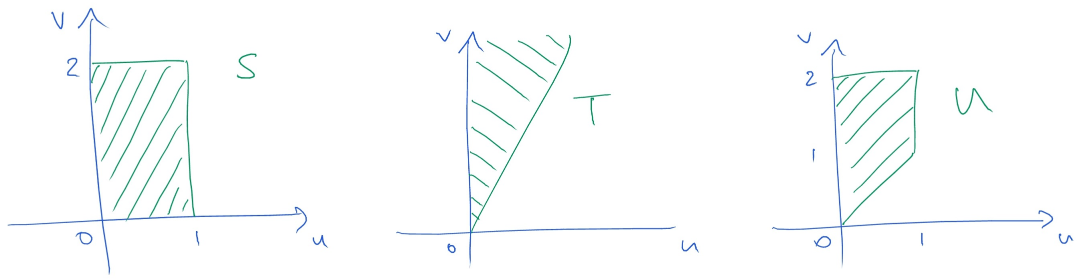
Warm-up Questions

5.1 (a) Define $u = x$ and $v = 2y$. Sketch the images of the regions S, T and U from question **4.2** in the (u, v) plane.

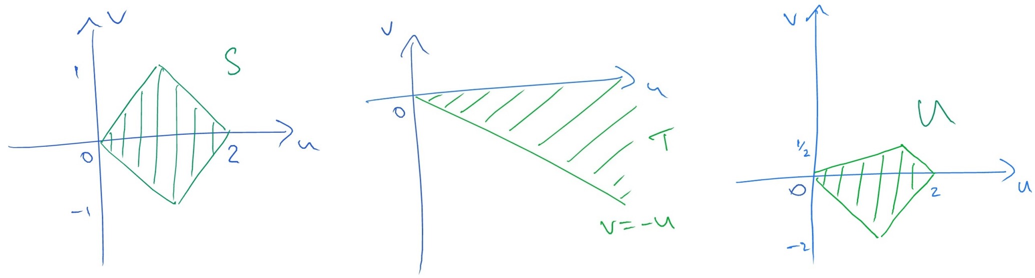
(b) Define $u = x + y$ and $v = x - y$. Sketch the images of the regions S, T and U from question **4.2** in the (u, v) plane.

Solution.

(a)



(b)



Ordinary Questions

5.2 The random variables X and Y have joint p.d.f. given by

$$f_{X,Y}(x, y) = \begin{cases} xe^{-y} & \text{if } x \in (0, 2), y \in (0, \infty) \\ 0 & \text{otherwise.} \end{cases}$$

Define $u = u(x, y) = x + y$ and $v = v(x, y) = 2y$. Let $U = u(X, Y)$ and $V = v(X, Y)$.

(a) Find the inverse transformation $x = x(u, v)$ and $y = y(u, v)$ and calculate the value of $J = \det \begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix}$.

(b) Sketch the set of (x, y) for which $f_{X,Y}(x, y)$ is non-zero. Find the image of this set in the (u, v) plane.

(c) Deduce from (a) and (b) that the joint p.d.f. of (U, V) is

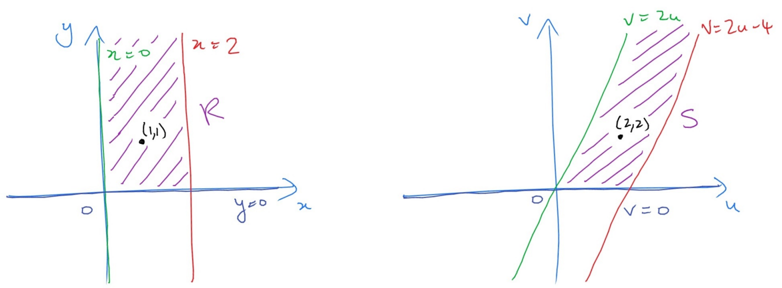
$$f_{U,V}(u,v) = \begin{cases} (u - \frac{v}{2})e^{-v/2} & \text{for } v > 0, u \in (2v, 2v+4) \\ 0 & \text{otherwise.} \end{cases}$$

Solution.

(a) The inverse transformation is $y = \frac{v}{2}$ and $x = u - y = u - \frac{v}{2}$. Hence,

$$J = \det \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix} = 1.$$

(b) $f_{X,Y}(x,y)$ is non-zero when $x \in (0,2)$ and $y \in (0, \infty)$. Call this set R. The set R is bounded by the three lines $x = 0$, $x = 2$ and $y = 0$. In the (u,v) plane these lines map respectively to $v = 2u$, $v = 2u - 4$ and $v = 0$. The point $(x,y) = (1,1) \in R$ maps to $(u,v) = (2,2)$, so R maps to the shaded region S in the (u,v) plane:



We can describe S as the set of (u,v) such that $v > 0$ and $v \in (2u-4, 2u)$.

(c) Using (a) and (b), we have

$$\begin{aligned} f_{U,V}(u,v) &= \begin{cases} f_{X,Y}\left(u - \frac{v}{2}, \frac{v}{2}\right) \times |J| & \text{for } (u,v) \in S \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} (u - \frac{v}{2})e^{-v/2} & \text{for } v > 0, v \in (2u-4, 2u) \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

5.3 The random variables X and Y have joint p.d.f. given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}(x+y)e^{-(x+y)} & \text{for } x, y \geq 0 \\ 0 & \text{elsewhere.} \end{cases}$$

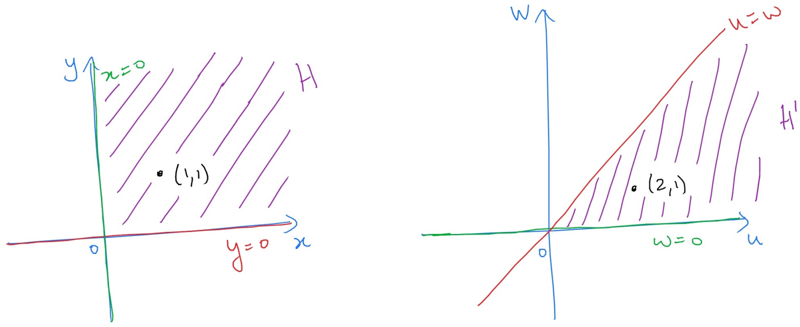
Let $U = X + Y$ and $W = X$.

(a) Find the joint p.d.f. of (U, W) and the marginal p.d.f. of U.

(b) Recognize U as a standard distribution and, using the result of Question **2.5**(a), evaluate $\mathbb{E}[(X+Y)^5]$.

Solution.

(a) Define $U = X + Y$ and $W = X$ so that $Y = U - X = U - W$. The set $H = \{(x, y) : x, y \geq 0\}$ is bounded by the lines $x = 0$ and $y = 0$, which respectively map to $w = 0$ and $u = w$. The point $(1, 1)$ is mapped to $(2, 1)$, so H is mapped to $H' = \{(u, w) : 0 \leq w \leq u\}$:



We have

$$\frac{\partial x}{\partial u} = 0, \quad \frac{\partial x}{\partial w} = 1, \quad \frac{\partial y}{\partial u} = 1, \quad \frac{\partial y}{\partial w} = -1$$

and so the Jacobian is

$$J = \det \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = -1.$$

We thus have

$$f_{U,W}(u, w) = f_{X,Y}(w, u - w) \times |J| = \begin{cases} \frac{1}{2} u e^{-u}, & 0 \leq w \leq u \\ 0, & \text{otherwise.} \end{cases}$$

Pitfall: We must remember to specify the region on which the pdf $f_{U,W}$ is non-zero; this is the region on which $f_{X,Y}$ is non-zero, mapped through the transformation $(x, y) \mapsto (u, v)$. The same applies to almost all questions in this section.

(b) Therefore, the marginal p.d.f. of U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,W}(u, w)\, dw = \int_0^u \frac{1}{2} u e^{-u}\, dw = \frac{1}{2} u e^{-u} [w]_0^u = \frac{1}{2} u^2 e^{-u} = \frac{1}{\Gamma(3)} u^{3-1} e^{-u}$$

for $u \geq 0$, and 0 otherwise, which means that $U \sim Ga(3, 1)$. From Question **2.5**, we have

$$\mathbb{E}[(X + Y)^k] = \mathbb{E}[U^k] = \frac{3(3 + 1) \cdots (3 + k - 1)}{1^k} = 3(4) \cdots (2 + k)$$

which, for $k = 5$, gives $\mathbb{E}[(X + Y)^5] = 3(4) \cdots (7) = \frac{7!}{2} = 2520$.

5.4 Let X and Y be a pair of independent random variables, both with the standard normal distribution. Show that the joint p.d.f. of (U, V) where $U = X^2$ and $V = X^2 + Y^2$ is given by

$$f_{U,V}(u, v) = \begin{cases} \frac{1}{8\pi} e^{-v/2} u^{-1/2} (v - u)^{-1/2} & \text{for } 0 \leq u \leq v \\ 0 & \text{otherwise.} \end{cases}$$

Solution. We have $X = \sqrt{U}$ and $Y = \sqrt{V - U}$. The density

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi} e^{-(x^2 + y^2)/2}$$

is positive for all $(x, y) \in \mathbb{R}^2$, and the transformation maps $\mathbb{R}^2$ to the set $\{(u, v) \,:\, 0 \le u \le v\}$. We have

$$\frac{\partial x}{\partial u} = \tfrac{1}{2}u^{-1/2}, \quad \frac{\partial x}{\partial v} = 0, \quad \frac{\partial y}{\partial u} = -\tfrac{1}{2}(v-u)^{-1/2} \quad \frac{\partial y}{\partial v} = \tfrac{1}{2}(v-u)^{-1/2}$$

and so the Jacobian is

$$J = \det \begin{pmatrix} \tfrac{1}{2}u^{-1/2} & 0 \\ -\tfrac{1}{2}(v-u)^{-1/2} & \tfrac{1}{2}(v-u)^{-1/2} \end{pmatrix} = \tfrac{1}{4}u^{-1/2}(v-u)^{-1/2}.$$

Hence, for $v \ge u \ge 0$ we have

$$f_{U,V}(u, v) = f_{X,Y}(\sqrt{u}, \sqrt{v-u}) \times |J| = \frac{1}{8\pi}e^{-v/2}u^{-1/2}(v-u)^{-1/2},$$

and $f_{U,V}(u, v) = 0$ otherwise.

5.5 Let (X, Y) be a random vector with joint p.d.f.

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-(x+y)} & x > y > 0; \\ 0 & \text{otherwise.} \end{cases}$$

(a) If $U = X - Y$ and $V = Y/2$, find the joint p.d.f. of (U, V).

(b) Show that U and V are independent, and recognize their (marginal) distributions as standard distributions.

Solution.

(a) We have

$$f_{X,Y} = \begin{cases} 2e^{-(x+y)} & x > y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and if $(u, v) = (x-y, y/2)$ then $(x, y) = (u+2v, 2v)$. The Jacobian of this transformation is

$$\det \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix} = 2.$$

The region $T = \{(x, y) \;:\; x > y > 0\}$ is described by $x > y$ and $y > 0$, which respectively become $u > 0$ and $v > 0$. So,

$$f_{U,V}(u, v) = \begin{cases} 4e^{-(u+4v)} & \text{for } u > 0, v > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(b) $f_{U,V}(u, v)$ factorises as $g(u)h(v)$ where

$$g(u) = \begin{cases} e^{-u} & u > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$h(v) = \begin{cases} 4e^{-4v} & v > 0 \\ 0 & \text{otherwise,} \end{cases}$$

so U and V are independent. We recognise $g(u)$ and $h(v)$ as the probability density functions of $Exp(1)$ and $Exp(4)$ random variables respectively, so $U \sim Exp(1)$ and $V \sim Exp(4)$.

Pitfall: We could calculate the marginal distributions of U and V by integrating out variables, but it is much more efficient to spot that we can factorize $f_{U,V}(u,v)$ into $f_U(u)f_V(v)$, where f_U and f_V are probability density functions. We know, in advance, that it will be possible to factorize in this way because they question told us that U and V would be independent.

The same issue also applies to many of the following questions.

5.6 Let X and Y be a pair of independent and identically distributed random variables. Let $U = X + Y$ and $V = X - Y$.

(a) Show that $\text{Cov}(U,V) = 0$, and give an example (with justification) to show that U and V are not necessarily independent.

(b) Show that U and V are independent in the special case where X and Y are standard normals.

Solution.

(a) We have

$$\begin{aligned}\text{Cov}(U,V) &= \mathbb{E}[(X+Y)(X-Y)] - \mathbb{E}[X+Y]\mathbb{E}[X-Y] \\ &= \mathbb{E}[X^2] - \mathbb{E}[Y^2] - \left(\mathbb{E}[X]^2 - \mathbb{E}[Y]^2\right) \\ &= \text{Var}(X) - \text{Var}(Y) = 0.\end{aligned}$$

The last line follows because X and Y have the same distribution. Hence, U and V are uncorrelated.

For the example, suppose that the common distribution of X and Y is such that they are equal to 1 with probability $\frac{1}{2}$ and equal to 0 with corresponding probability $\frac{1}{2}$. Then, $\mathbb{P}[U=2] = \frac{1}{4}$, and if $U = 2$ then $X = Y = 1$, which means $V = 0$. So U and V are not independent;

$$\mathbb{P}[U=2, V=0] = \frac{1}{4} \quad \neq \quad \mathbb{P}[U=2]\mathbb{P}[V=0] = \mathbb{P}[X=Y=1]\mathbb{P}[X=Y=0] = \frac{1}{16}.$$

In fact, most possible examples result in U and V not being independent, and this one is chosen for its simplicity.

(b) If X and Y are standard normals, then by independence we have

$$f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}.$$

The transformation $U = X + Y$, $V = X - Y$ maps $\mathbb{R}^2$ to $\mathbb{R}^2$, and has inverse transformation $X = \frac{1}{2}(U+V)$, $Y = \frac{1}{2}(U-V)$. We have

$$\frac{\partial x}{\partial u} = \frac{1}{2}, \quad \frac{\partial x}{\partial v} = \frac{1}{2}, \quad \frac{\partial y}{\partial u} = \frac{1}{2}, \quad \frac{\partial y}{\partial v} = \frac{-1}{2}$$

and hence the Jacobian is

$$J = \det\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{-1}{2} \end{pmatrix} = -\frac{1}{2}.$$

Hence, the p.d.f. of (U,V) is given by

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}\left(\tfrac{1}{2}(u+v), \tfrac{1}{2}(u-v)\right) \times |J| \\ &= \frac{1}{4\pi}\exp\left(-\tfrac{1}{8}\left((u+v)^2+(u-v)^2\right)\right) \\ &= \frac{1}{4\pi}e^{-(u^2+v^2)/4} \\ &= \left(\frac{1}{\sqrt{4\pi}}e^{-u^2/4}\right)\left(\frac{1}{\sqrt{4\pi}}e^{-v^2/4}\right). \end{aligned}$$

Hence, U and V are independent, and both U and V have the $N(0,2)$ distribution.

5.7 Let X and Y be independent random variables with distributions $Ga(\alpha_1,\beta)$ and $Ga(\alpha_2,\beta)$ respectively. Show that the random variables $U=\frac{X}{X+Y}$ and $V=X+Y$ are independent with distributions $Be(\alpha_1,\alpha_2)$ and $Ga(\alpha_1+\alpha_2,\beta)$ respectively.

Solution. Since X and Y are independent we have

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_Y(y) = \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)}x^{\alpha_1-1}e^{-\beta x}\frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)}y^{\alpha_2-1}e^{-\beta y} \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}x^{\alpha_1-1}y^{\alpha_2-1}e^{-\beta(x+y)}. \end{aligned} \tag{5.1}$$

If $v=x+y$ and $u=\frac{x}{x+y}$ then, to find the inverse transformation we observe that $uv=(x+y)\frac{x}{x+y}=x$, so $x=uv$ and hence $y=v-uv$. The conditions $x>0$ and $y>0$ translate to $v>0$ and $0<u<1$. The partial derivatives are

$$\frac{\partial x}{\partial u}=v,\quad \frac{\partial x}{\partial v}=u,\quad \frac{\partial y}{\partial u}=-v,\quad \frac{\partial y}{\partial v}=1-u$$

and the Jacobian is

$$J=\det\begin{pmatrix} v & u \\ -v & 1-u\end{pmatrix} = v(1-u)+uv = v-uv+uv = v.$$

Hence, for $v>0$ and $u\in(0,1)$, the joint p.d.f. of U and V is

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(uv, v-uv)v = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}(uv)^{\alpha_1-1}(v-uv)^{\alpha_2-1}e^{-\beta v}v \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}u^{\alpha_1-1}(1-u)^{\alpha_2-1}v^{\alpha_1+\alpha_2-1}e^{-\beta v}. \end{aligned}$$

For all other u,v, we have $f_{U,V}(u,v)=0$.

At this point we can recall that $\beta(\alpha_1,\alpha_2)=\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$ and spot that $f_{U,V}(u,v)$ factorises as

$$f_{U,V}(u,v)=\begin{cases} \frac{1}{B(\alpha_1,\alpha_2)}u^{\alpha_1-1}(1-u)^{\alpha_2-1}\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}v^{\alpha_1+\alpha_2-1}e^{-\beta v} & \text{for } v>0, u\in(0,1) \\ 0 & \text{otherwise.}\end{cases}$$

It follows immediately that $V\sim Ga(\alpha_1+\alpha_2,\beta)$ and that U and V are independent.

5.8 As part of Question **5.7**, we showed that if X and Y are independent random variables with $X\sim Ga(\alpha_1,\beta)$ and $Y\sim Ga(\alpha_2,\beta)$, then $X+Y\sim Ga(\alpha_1+\alpha_2,\beta)$.

(a) Use induction to show that for $n \geq 2$, if $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i \sim Ga(\alpha_i, \beta)$ then

$$\sum_{i=1}^{n} X_i \sim Ga\left(\sum_{i=1}^{n} \alpha_i, \beta\right).$$

(b) Hence show that for $n \geq 1$, if $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables then

$$\sum_{i=1}^{n} Z_i^2 \sim \chi_n^2.$$

You may use the result of Example 12, which showed that this was true in the case $n = 1$ (and, recall that the χ^2 distribution is a special case of the Gamma distribution).

Solution.

(a) The base case $n = 2$ is the given result from Question **5.7**. If the claim is true for $n = k$, then let $X = \sum_{i=1}^{k} X_i$ and $Y \sim X_{k+1}$, so that $X \sim Ga\left(\sum_{i=1}^{k} \alpha_i, \beta\right)$ and $Y \sim Ga(\alpha_{k+1}, \beta)$. Then the same result from Question **5.7** shows that

$$\sum_{i=1}^{k+1} X_i = X + Y \sim Ga\left(\sum_{i=1}^{k+1} \alpha_i, \beta\right),$$

so the claim is true for $n = k + 1$. Hence it is true for all $n \geq 2$ by induction.

(b) Recall that the χ_n^2 distribution is the same as the $Ga\left(\frac{n}{2}, \frac{1}{2}\right)$ distribution. In particular each $Z_i^2 \sim Ga\left(\frac{1}{2}, \frac{1}{2}\right)$, so the result of (a) tells us that $\sum_{i=1}^{n} Z_i^2 \sim Ga\left(\frac{n}{2}, \frac{1}{2}\right) = \chi_n^2$ as required.

5.9 Let X and Y be a pair of independent random variables, both with the standard normal distribution. Show that $U = X/Y$ has the Cauchy distribution, with p.d.f.

$$f_U(u) = \frac{1}{\pi} \frac{1}{1 + u^2}$$

for all $u \in \mathbb{R}$.

Solution. The joint density

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

is positive for all $(x, y) \in \mathbb{R}^2$. We define $U = X/Y$ and $V = X$ (which makes sense because $\mathbb{P}[Y = 0] = 0$). We set $u = x/y$ and $v = x$, and note that this transformation maps $\mathbb{R}^2$ to $\mathbb{R}^2$. The inverse transformation is $x = v$ and $y = v/u$, which gives

$$\frac{\partial x}{\partial u} = 0, \quad \frac{\partial x}{\partial v} = 1, \quad \frac{\partial y}{\partial u} = \frac{-v}{u^2}, \quad \frac{\partial y}{\partial v} = \frac{1}{u},$$

and hence the Jacobian is

$$J = \det \begin{pmatrix} 0 & 1 \\ \frac{-v}{u^2} & \frac{1}{u} \end{pmatrix} = \frac{v}{u^2}.$$

Hence,

$$f_{U,V}(u,v) = f_{X,Y}(v, v/u) \times |J| = \frac{1}{2\pi} \exp\left(-\frac{v^2(1+1/u^2)}{2}\right) \frac{|v|}{u^2}$$

for all $u, v \in \mathbb{R}^2$. To obtain $f_U(u)$ we integrate out v, giving

$$\begin{aligned}
f_U(u) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|v|}{u^2} \exp\left(-\frac{v^2(1+1/u^2)}{2}\right) dv \\
&= \frac{1}{2\pi} 2 \int_0^{\infty} \frac{v}{u^2} \exp\left(-\frac{v^2(1+1/u^2)}{2}\right) dv \\
&= \frac{1}{\pi} \frac{1}{u^2} \frac{1}{1+1/u^2} \int_0^{\infty} v(1+1/u^2) \exp\left(-\frac{v^2(1+1/u^2)}{2}\right) dv \\
&= \frac{1}{\pi} \frac{1}{1+u^2} \left[-\exp\left(-\frac{v^2(1+1/u^2)}{2}\right)\right]_{v=0}^{\infty} \\
&= \frac{1}{\pi} \frac{1}{1+u^2}.
\end{aligned}$$

This matches the p.d.f. of the Cauchy distribution.

5.10 Let $n \in \mathbb{N}$. The t distribution (often known as Student's t distribution) is the univariate random variable X with p.d.f.

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

for all $x \in \mathbb{R}$. Here, n is a parameter, known as the number of degrees of freedom.

Let Z be a standard normal random variable and let W be a chi-squared random variable with n degrees of freedom, where Z and W are independent. Show that

$$X = \frac{Z}{\sqrt{W/n}}$$

has the t distribution with n degrees of freedom.

Solution. To find the probability density function of X, consider (Z, W) as a bivariate random vector, and transform it to (X, Y) where $X = \frac{Z}{\sqrt{W/n}}$ as above and $Y = W$.

By independence and the formulae for the density functions of the Normal and chi-squared distributions,

$$f_{Z,W}(z,w) = \frac{1}{\sqrt{2^n}\Gamma(n/2)\sqrt{2\pi}} w^{\frac{n}{2}-1} \exp\left(-\frac{(z^2+w)}{2}\right),$$

for $w > 0$. We set $x = z/\sqrt{w/n}$ and $y = w$, and note that the inverse of this transformation is given by $w = y$ and $z = x\sqrt{y/n}$. Moreover, this transformation maps $\{(z,w) \,:\, z \in \mathbb{R}, w > 0\}$ to $\{(x,y) \,:\, x \in \mathbb{R}, y > 0\}$. The Jacobian is

$$\det \begin{pmatrix} 0 & 1 \\ \sqrt{y/n} & x(y/n)^{1/2}/(2n) \end{pmatrix} = -\sqrt{y/n},$$

so the joint p.d.f. of X and Y is

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{\sqrt{2^n}\Gamma(n/2)\sqrt{2\pi}} y^{\frac{n}{2}-1} \exp\left(-\frac{\left(\frac{x^2 y}{n}+y\right)}{2}\right)\sqrt{y/n} \\ &= \frac{1}{\sqrt{2^n}\Gamma(n/2)\sqrt{2\pi n}} y^{\frac{n-1}{2}} \exp\left(-\frac{y}{2}\left(\frac{x^2}{n}+1\right)\right), \end{aligned}$$

when $y > 0$, and equal to zero when $y \le 0$.

To obtain the p.d.f. of X, we integrate out y:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\ &= \frac{1}{\sqrt{2^n}\Gamma(n/2)\sqrt{2\pi n}} \int_0^\infty y^{\frac{n-1}{2}} \exp\left(-\frac{y}{2}\left(\frac{x^2}{n}+1\right)\right)\,dy. \end{aligned}$$

By Lemma 2.3,

$$\int_0^\infty y^{\frac{n-1}{2}} \exp\left(-\frac{y}{2}\left(\frac{x^2}{n}+1\right)\right)\,dy = \left(\frac{1}{2}\left(\frac{x^2}{n}+1\right)\right)^{-(n+1)/2} \Gamma\left(\frac{n+1}{2}\right),$$

giving

$$f_X(x) = \frac{1}{\sqrt{2^n}\Gamma(n/2)\sqrt{2\pi n}} \left(\frac{1}{2}\left(\frac{x^2}{n}+1\right)\right)^{-(n+1)/2} \Gamma\left(\frac{n+1}{2}\right),$$

which simplifies to

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

as required.

Challenge Questions

5.11 The formula (5.2), from the typed lecture notes, holds whenever the transformation $u = u(x,y)$, $v = v(x,y)$ is both one-to-one and onto, and for which the Jacobian matrix of derivatives exists. Use this fact to provide an alternative proof of Lemma 3.1 (i.e. of the univariate transformation formula).

Hint. Take a random variable X and a function g that satisfies the conditions of Lemma 3.1. Let W be a continuous random variable that is independent of X. Define $y = g(x)$, $z = w$, and consider the pair $(Y,Z) = (g(X), W)$. Find $f_{Y,Z}(y,z)$, using equation (5.2), and integrate out z to obtain $f_Y(y)$.

5.12 Let $X \sim Exp(\lambda_1)$ and $Y \sim Exp(\lambda_2)$ be independent. Show that $U = \min(X,Y)$ has distribution $Exp(\lambda_1+\lambda_2)$, and that $\mathbb{P}[\min(X,Y) = X] = \frac{\lambda_1}{\lambda_1+\lambda_2}$. Extend this result, by induction, to handle the minimum of finitely many exponential random variables.

Let $W = \max(X,Y)$. Show that U and $W-U$ are independent.

Hint. Calculate $\mathbb{P}[\min(X,Y) \le z]$ directly by partitioning on the event $X < Y$.

6 Covariance Matrices and and Multivariate Normal Distributions

Warm-up Questions

6.1 Let

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix}.$$

(a) Calculate $\det(\mathbf{A})$ and find both $\mathbf{A\Sigma}$ and $\mathbf{A\Sigma A}^T$.

(b) Let $\mathbf{A}$ be the vector $(2, 1)$. Show that $\mathbf{A\Sigma A}^T = 21$.

Solution.

(a) We have $\det(A) = (1)(1) - (-1)(2) = 3$, and

$$\begin{aligned} \mathbf{A\Sigma} &= \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix} = \begin{pmatrix} 5 & -10 \\ 7 & 7 \end{pmatrix} \\ \mathbf{A\Sigma A}^T &= \begin{pmatrix} 5 & -10 \\ 7 & 7 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 15 & 0 \\ 0 & 21 \end{pmatrix}. \end{aligned}$$

Alternatively, we could calculate $\mathbf{\Sigma A}^T$ first and then pre-multiply by $\mathbf{A}$.

(b) We have $\mathbf{A\Sigma A}^T = \begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 7 \\ 7 \end{pmatrix} = 21.$

Ordinary Questions

6.2 Let $\mathbf{X} = (X, Y)^T$ be a random vector with

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \mathrm{Cov}(\mathbf{X}) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Let $U = X + Y$ and $V = 2X - 2Y + 1$. Write $\mathbf{U} = (U, V)^T$.

(a) Write down a square matrix $\mathbf{A}$ and a vector $\mathbf{b} \in \mathbb{R}^2$ such that $\mathbf{U} = \mathbf{AX} + \mathbf{b}$.

(b) Show that the mean vector and covariance matrix of $\mathbf{U}$ are given by

$$\mathbb{E}[\mathbf{U}] = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \mathrm{Cov}(\mathbf{U}) = \begin{pmatrix} 5 & -2 \\ -2 & 4 \end{pmatrix}$$

(c) Show that the correlation coefficient ρ of U and V is equal to $\frac{-1}{\sqrt{5}}$.

Solution.

(a) We have

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

so we take $\mathbf{A} = \left(\begin{smallmatrix} 1 & 1 \\ 2 & -2 \end{smallmatrix}\right)$ and $\mathbf{b} = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$.

(b) Using Lemma 6.3 we have

$$\begin{aligned}\mathbb{E}[\mathbf{U}] &= \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b} \\ &= \begin{pmatrix}1 & 1\\ 2 & -2\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\-2\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix}\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(\mathbf{U}) &= \mathbf{A}\,\mathrm{Cov}(\mathbf{X})\mathbf{A}^T \\ &= \begin{pmatrix}1 & 1\\ 2 & -2\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}1 & 2\\ 1 & -2\end{pmatrix} = \begin{pmatrix}1 & 1\\ 2 & -2\end{pmatrix}\begin{pmatrix}2 & 0\\ 3 & -2\end{pmatrix} = \begin{pmatrix}5 & -2\\ -2 & 4\end{pmatrix}\end{aligned}$$

(c) From the covariance matrix of $\mathbf{U}$,

$$\rho(U,V) = \frac{\mathrm{Cov}(U,V)}{\sqrt{\mathrm{Var}(U)\,\mathrm{Var}(V)}} = \frac{-2}{\sqrt{5\times 4}} = \frac{-1}{\sqrt{5}}.$$

6.3 Let X and Y be independent standard normal random variables.

(a) Write down the covariance matrix of the random vector $\mathbf{X} = (X,Y)^T$.

(b) Let $\mathbf{R}$ be the rotation matrix

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

Show that $\mathbf{RX}$ has the same covariance matrix as $\mathbf{X}$.

Solution.

(a) Since X and Y are standard normals they both have variance 1, and by independence their covariance is zero. Hence $\mathrm{Cov}(\mathbf{X}) = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. This is the identify matrix, which we denote by I.

(b) We have $\mathrm{Cov}(\mathbf{RX}) = \mathbf{R}\,\mathrm{Cov}(\mathbf{X})\mathbf{R}^T = \mathbf{R}I\mathbf{R}^T = \mathbf{RR}^T$, and evaluating $\mathbf{RR}^T$ gives

$$\begin{aligned}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} &= \begin{pmatrix}\cos^2\theta + \sin^2\theta & \cos\theta\sin\theta - \sin\theta\cos\theta\\ \sin\theta\cos\theta - \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta\end{pmatrix} \\ &= \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\end{aligned}$$

which is the identity matrix I. Since $\mathrm{Cov}(\mathbf{X}) = I$, we are done.

6.4 Three (univariate) random variables X, Y and Z have means 3, -4 and 6 respectively and variances 1, 1 and 25 respectively. Further, X and Y are uncorrelated; the correlation coefficient between X and Z is $\frac{1}{5}$ and that between Y and Z is $-\frac{1}{5}$. Let $U = X + Y - Z$ and $W = 2X + Z - 4$ and set $\mathbf{U} = (U,W)^T$.

(a) Find the mean vector and covariance matrix of $\mathbf{X} = (X,Y,Z)^T$.

(b) Write down a matrix $\mathbf{A}$ and a vector $\mathbf{b}$ such that $\mathbf{U} = \mathbf{AX} + \mathbf{b}$.

(c) Find the mean vector and covariance matrix of $\mathbf{U}$.

(d) Evaluate $\mathbb{E}[(2X+Z-6)^2]$.

Solution.

(a) We have $\mathbb{E}[(X,Y,Z)^T] = (\mathbb{E}[X],\mathbb{E}[Y],\mathbb{E}[Z])^T = (3,-4,6)^T$. For the covariance matrix, we have $\mathrm{Cov}(X,Y)=0$ since X and Y are uncorrelated. We are given that $\rho_{X,Z}=\frac{1}{5}$, so $\frac{\mathrm{Cov}(X,Z)}{\sqrt{1\times 25}}=\frac{1}{5}$ and hence $\mathrm{Cov}(X,Z)=1$. Similarly $\rho_{Y,Z}=-\frac{1}{5}$ gives $\mathrm{Cov}(X,Z)=-1$. So,

$$\mathrm{Cov}\begin{pmatrix}X\\Y\\Z\end{pmatrix} = \begin{pmatrix}1&0&1\\0&1&-1\\1&-1&25\end{pmatrix}.$$

(b) We define $\mathbf{A}$ and $\mathbf{b}$ by

$$\begin{pmatrix}U\\W\end{pmatrix} = \begin{pmatrix}1&1&-1\\2&0&1\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix} + \begin{pmatrix}0\\-4\end{pmatrix} = \mathbf{AX}+\mathbf{b}$$

(c) It holds that

$$\mathbb{E}\begin{pmatrix}U\\W\end{pmatrix} = \mathbf{A}\mathbb{E}[\mathbf{X}]+\mathbf{b} = \begin{pmatrix}1&1&-1\\2&0&1\end{pmatrix}\begin{pmatrix}3\\-4\\6\end{pmatrix} + \begin{pmatrix}0\\-4\end{pmatrix} = \begin{pmatrix}-7\\8\end{pmatrix}$$

and

$$\mathrm{Cov}\begin{pmatrix}U\\W\end{pmatrix} = \mathbf{A}\,\mathrm{Cov}(\mathbf{X})\mathbf{A}^T = \begin{pmatrix}27&-25\\-25&33\end{pmatrix}.$$

(d) We observe $\mathbb{E}[(2X+Z-6)^2]=\mathbb{E}[(W-2)^2]$. From (b) we have

$$\mathrm{Var}(W-2)=\mathrm{Var}(W)=33$$

and $\mathbb{E}[W]=8$, so $\mathbb{E}[W-2]=6$. Hence, by rearranging $\mathrm{Var}(W-2)=\mathbb{E}[(W-2)^2]-\mathbb{E}[W-2]^2$,

$$\mathbb{E}[(W-2)^2]=\mathrm{Var}(W-2)+\mathbb{E}[W-2]^2 = 33+6^2=69.$$

6.5 Suppose that the random vector $\mathbf{X}=(X_1,X_2)^T$ follows the bivariate normal distribution with $\mathbb{E}[X_1]=\mathbb{E}[X_2]=0$, $\mathrm{Var}(X_1)=1$, $\mathrm{Cov}(X_1,X_2)=2$ and $\mathrm{Var}(X_2)=5$.

(a) Calculate the correlation coefficient of X_1 and X_2. Are X_1 and X_2 independent?

(b) Find the mean and the covariance matrices of

$$Y=\begin{pmatrix}1&2\end{pmatrix}\mathbf{X} \qquad \text{and} \qquad \mathbf{Z}=\begin{pmatrix}1&2\\3&4\end{pmatrix}\mathbf{X}.$$

What are the distributions of Y and $\mathbf{Z}$?

Solution.

(a) We have

$$\rho = \frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{\mathrm{var}(X_1)\mathrm{var}(X_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{2}{1\times\sqrt{5}} \approx 0.894.$$

X_1 and X_2 are positively correlated and they are not independent.

(b) We have

$$\mathbb{E}[Y] = (1,2)\mathbb{E}[\mathbf{X}] = (1,2)\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

$$\mathrm{Cov}(Y) = (1,2)\,\mathrm{Cov}(\mathbf{X})\begin{pmatrix} 1 \\ 2 \end{pmatrix} = (1,2)\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = (5,12)\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 29.$$

and

$$\mathbb{E}[\mathbf{Z}] = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\mathbb{E}[\mathbf{X}] = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$$\begin{aligned}\mathrm{Cov}(\mathbf{Z}) &= \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\mathrm{Cov}(\mathbf{X})\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T \\ &= \begin{pmatrix} 5 & 12 \\ 11 & 26 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 29 & 63 \\ 63 & 137 \end{pmatrix}.\end{aligned}$$

In the cases of both Y and $\mathbf{Z}$, the transformations which obtain them from $\mathbf{X}$ are linear. Hence, both Y and $\mathbf{Z}$ are normally distributed. Because the normal distribution is determined completely from its mean and its covariance matrix, it follows that the distributions of Y and $\mathbf{Z}$ are

$$Y \sim N(0,29) \quad \text{and} \quad \mathbf{Z} \sim N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 29 & 63 \\ 63 & 137 \end{pmatrix}\right].$$

6.6 Let X_1 and X_2 be bivariate normally distributed random variables each with mean 0 and variance 1, and with correlation coefficient ρ.

(a) By integrating out the variable x_2 in the joint p.d.f., verify that the marginal distribution of X_1 is indeed that of a standard univariate normal random variable. *Hint: Use the fact that the integral of a $N(\mu, \sigma^2)$ p.d.f. is equal to* 1.

(b) Show, using the 'usual' formula for the conditional p.d.f. that the conditional p.d.f. of X_2 given $X_1 = x_1$ is $N(\rho x_1, 1-\rho^2)$.

Solution.

(a) We have

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[x_1^2 - 2\rho x_1 x_2 + x_2^2\right]\right\},$$

so integrating out x_2 gives

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{\left[x_1^2 - 2\rho x_1 x_2 + x_2^2\right]}{2(1-\rho^2)}\right\}dx_2 \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{\left[(x_2-\rho x_1)^2 - (\rho x_1)^2 + x_1^2\right]}{2(1-\rho^2)}\right\}dx_2 \\ &\quad \text{(by completing the square)} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{\frac{-x_1^2(1-\rho^2)}{2(1-\rho^2)}\right\}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x_2-\rho x_1)^2}{2(1-\rho^2)}\right\}dx_2 \\ &\quad \text{(by taking terms which don't involve } x_2 \text{ outside the integral)} \\ &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x_1^2}{2}\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left\{\frac{-(x_2-\rho x_1)^2}{2(1-\rho^2)}\right\}dx_2. \end{aligned}$$

The term inside the integral here is the p.d.f. of a Normal random variable with mean ρx_1 and variance $(1-\rho^2)$. (Because the integral is with respect to x_2, we can treat x_1 as a constant). Hence, the integral is 1, and so we have shown

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}},$$

which is the p.d.f. of the standard normal.

(b) From (a), we have

$$\begin{aligned}
f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} \\
&= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[x_1^2 - 2\rho x_1 x_2 + x_2^2\right]\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right)} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}x_1^2 + \frac{2\rho}{2(1-\rho^2)}x_1x_2 - \frac{1}{2(1-\rho^2)}x_2^2 + \frac{x_1^2}{2}\right] \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2}\left[\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2 - (1-\rho^2)x_1^2}{1-\rho^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(x_2 - \rho x_1)^2}{2(1-\rho^2)}\right]
\end{aligned}$$

and so $X_2|X_1 = x_1 \sim N(\rho x_1, 1-\rho^2)$ as required.

6.7 Let $\mathbf{X} = (X_1, X_2)^T$ have a $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$. Let

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{-\sigma_1}{\sigma_2} \\ \frac{\sigma_2}{\sigma_1} & 1 \end{pmatrix}.$$

Find the distribution of $\mathbf{Y} = \mathbf{AX}$, and deduce that any bivariate normal random vector can be transformed by a linear transformation into a vector of independent normal random variables.

Solution. The random vector $\mathbf{Y}$ has a bivariate normal distribution with mean vector

$$\mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} \mu_1 - \frac{\mu_2\sigma_1}{\sigma_2} \\ \mu_2 + \frac{\mu_1\sigma_2}{\sigma_1} \end{pmatrix},$$

and covariance matrix

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T = \begin{pmatrix} 1 & -\frac{\sigma_1}{\sigma_2} \\ \frac{\sigma_2}{\sigma_1} & 1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & \frac{\sigma_2}{\sigma_1} \\ -\frac{\sigma_1}{\sigma_2} & 1 \end{pmatrix} = \begin{pmatrix} \frac{2\sigma_1(\sigma_1\sigma_2 - \sigma_{12})}{\sigma_2} & 0 \\ 0 & \frac{2\sigma_2(\sigma_1\sigma_2 + \sigma_{12})}{\sigma_1} \end{pmatrix}.$$

So

$$\mathbf{Y} \sim N_2\left[\begin{pmatrix} \mu_1 - \frac{\mu_2\sigma_1}{\sigma_2} \\ \mu_2 + \frac{\mu_1\sigma_2}{\sigma_1} \end{pmatrix}, \begin{pmatrix} \frac{2\sigma_1(\sigma_1\sigma_2 - \sigma_{12})}{\sigma_2} & 0 \\ 0 & \frac{2\sigma_2(\sigma_1\sigma_2 + \sigma_{12})}{\sigma_1} \end{pmatrix}\right].$$

Hence, the components of $\mathbf{Y}$ are independent.

Since this transformation can be applied to any bivariate normal random vector, any bivariate normal random vector can be transformed (by a linear transformation) into a vector of independent normal random variables.

6.8 The random vector $\mathbf{X} = (X_1, X_2, X_3)^T$ has an $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution where $\boldsymbol{\mu} = (-1, 1, 2)^T$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 144 & -30 & 48 \\ -30 & 25 & 10 \\ 48 & 10 & 64 \end{pmatrix}.$$

(a) Find the correlation coefficients between X_1 and X_2, between X_1 and X_3 and between X_2 and X_3.

(b) Let $Y_1 = X_1 + X_3$ and $Y_2 = X_2 - X_1$. Find the distribution of $\mathbf{Y} = (Y_1, Y_2)^T$ and hence find the correlation coefficient between Y_1 and Y_2.

Solution.

(a) Using that $\rho(X, Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$, from the covariance matrix, the correlation coefficient between X_1 and X_2 is $\frac{-30}{\sqrt{144}\sqrt{25}} = -\frac{1}{2}$, that between X_1 and X_3 is $\frac{48}{\sqrt{144}\sqrt{64}} = \frac{1}{2}$ and that between X_2 and X_3 is $\frac{10}{\sqrt{25}\sqrt{64}} = \frac{1}{4}$.

(b) We have $\mathbf{Y} = \mathbf{AX}$ where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}.$$

The mean vector of $\mathbf{Y}$ is $\mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, and its covariance matrix is given by

$$\mathbf{A}\Sigma\mathbf{A}^T = \begin{pmatrix} 304 & -212 \\ -212 & 229 \end{pmatrix}.$$

Hence, the correlation between Y_1 and Y_2 is $\frac{-212}{\sqrt{304}\sqrt{229}} \approx -0.803$.

6.9 (a) Let X be a (univariate) standard normal random variable and let $Y = X$. Does the random vector $\mathbf{X} = (X, Y)^T$ have a bivariate normal distribution?

(b) Let X and Y be two independent (univariate) standard normal random variables, and let Z be a random variable such that $\mathbb{P}[Z = 1] = \mathbb{P}[Z = -1] = \frac{1}{2}$. Are XZ and YZ independent, and does the random vector $\mathbf{X} = (XZ, YZ)^T$ have a bivariate normal distribution?

Solution.

(a) No. To see this, for example, note that a bivariate normal $\mathbf{Z} = (Z_1, Z_2)^T$ has a p.d.f. $f_{Z_1,Z_2}(z_1, z_2)$ that is positive everywhere, and so

$$\mathbb{P}[Z_1 < Z_2] = \int_{-\infty}^{\infty} \int_{z_1}^{\infty} f_{Z_1,Z_2}(z_1, z_2)\, dz_2\, dz_1 > 0,$$

but $\mathbb{P}[X < Y] = 0$.

Another possible way to see it: note that $\operatorname{Cov}(X, Y) = \operatorname{Cov}(X, X) = \operatorname{Var}(X) = 1$, but if X and Y were independent they would have covariance zero.

(b) Yes. To see this, note that for any subsets A, B of $\mathbb{R}$ we have

$$\begin{aligned}\mathbb{P}[XZ \in A, YZ \in B] &= \mathbb{P}[X \in A, Y \in B, Z = 1] + \mathbb{P}[-X \in A, -Y \in B, Z = -1] \\ &= \mathbb{P}[X \in A, Y \in B]\mathbb{P}[Z = 1] + \mathbb{P}[-X \in A, -Y \in B]\mathbb{P}[Z = -1] \\ &= \mathbb{P}[X \in A, Y \in B]\tfrac{1}{2} + \mathbb{P}[X \in A, Y \in B]\tfrac{1}{2} \\ &= \mathbb{P}[X \in A, Y \in B].\end{aligned}$$

Hence $(XZ, YZ)^T$ has the same distribution as $(X, Y)^T$, which is that of a pair of independent normal random variables. Hence (X, Y) has a bivariate normal distribution, with mean vector $\boldsymbol{\mu} = \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$ and covariance matrix $\boldsymbol{\Sigma} = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$.

Challenge Questions

6.10 Recall that an *orthogonal matrix* $\mathbf{R}$ is one for which $\mathbf{R}^{-1} = \mathbf{R}^T$, and recall that if an $n \times n$ matrix $\mathbf{R}$ is orthogonal, $\mathbf{x}$ is an n-dimensional vector, and $\mathbf{y} = \mathbf{R}\mathbf{x}$ then $\sum_{i=1}^n y_i^2 = \mathbf{y} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^n x_i^2$.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of independent normal random variables with common mean 0 and variance σ^2. Let $\mathbf{R}$ be an orthogonal matrix and let $\mathbf{Y} = \mathbf{R}\mathbf{X}$.

(a) Show that $\mathbf{Y}$ is also a vector of independent normal random variables, with common mean 0 and variance σ^2.

(b) Suppose that all the elements in the first row of $\mathbf{R}$ are equal to $\frac{1}{\sqrt{n}}$ (you may assume that an orthogonal matrix exists with this property). Show that $Y_1 = \sqrt{n}\bar{X}$, where $\bar{X}$ is the sample mean of $\mathbf{X}$, and that

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2.$$

(c) Hence, use Question **5.8** to deduce that, if $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)$ is the sample variance of $\mathbf{X}$,

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$

and that it is independent of $\bar{X}$.

(d) Let $\mu \in \mathbb{R}$. Deduce that the result of part (c) also holds if the X_i have mean μ (so as they are i.i.d. $N(\mu, \sigma^2)$ random variables).

Solution.

(a) We note that $\mathbf{X} \sim N_n(\mathbf{0}, \sigma^2 I)$, where $\mathbf{0}$ is an n-dimensional vector of zeros and I is the n-dimensional identity matrix. The mean vector of $\mathbf{Y}$ will then be $\mathbf{R0} = \mathbf{0}$, and the covariance matrix will be $R(\sigma^2 I)R^T = \sigma^2\mathbf{R}\mathbf{R}^T = \sigma^2 I$, by the orthogonality of $\mathbf{R}$, so the multivariate normal theory tells us that $\mathbf{Y}$ also has $N_n(\mathbf{0}, \sigma^2 I)$ distribution.
*Note that this result is a generalisation of **6.3**, since* $\left(\begin{smallmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{smallmatrix}\right)$ *is orthogonal.*

(b) By the form given for **R**,

$$Y_1 = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} X_i = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} X_i = \sqrt{n}\bar{X}.$$

By the orthogonality of **R** we have $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2$, so

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2 - Y_1^2 = \sum_{i=1}^{n} X_i^2 - n(\bar{X})^2.$$

(c) By part (a), the Y_i are independent $N(0, \sigma^2)$ random variables, so $\frac{Y_i}{\sigma} \sim N(0,1)$, and by **5.8**

$$\sum_{i=2}^{n} \left(\frac{Y_i}{\sigma}\right)^2 \sim \chi^2_{n-1}.$$

Hence by part (b) we have

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \sim \chi^2_{n-1},$$

as required. The independence follows because Y_1 is independent of the remaining variables by the form of the covariance matrix and multivariate normal theory.

(d) Recall that an alternate for of the sample variance of (X_i) is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Let $X_i' = X_i + \mu$, so as the X_i' are i.i.d. $N(\mu, \sigma^2)$. Write $\bar{X}' = \frac{1}{n} \sum_{i=1}^{n} X_i' = \bar{X} + \mu$. Hence,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left((X_i' - \mu) - (\bar{X}' - \mu)\right)^2,$$

that is the (X_i) and the (X_i') have the same sample variance. So, the sample variance of the (X_i') also has the χ^2_{n-1} distribution.

6.11 Let $Z_1, Z_2, \ldots$ be independent identically distributed normal random variables with mean μ and variance σ^2. We regard n samples of these as a set of data. Write $\bar{Z}$ and s^2 respectively for the sample mean and variance.

Combine **6.10**(d) and **5.10** to show that the statistic

$$X' = \frac{\sqrt{n}(\bar{Z} - \mu)}{s}$$

has the t distribution with $n-1$ degrees of freedom.

Solution. We note,

$$X' = \frac{\frac{\sqrt{n}}{\sigma}\left(\bar{Z} - \mu\right)}{\sqrt{\frac{(n-1)s^2}{\sigma^2}}\sqrt{\frac{1}{n-1}}}.$$

This rearrangement (writing $s = \sqrt{s^2}$, then multiplying and dividing appropriately by σ and $n-1$), is chosen precisely to create the term $W = \sqrt{\frac{(n-1)s^2}{\sigma^2}}$ on the bottom. By **6.10**(d), W has a χ^2_{n-1} distribution, independent of the numerator $\frac{\sqrt{n}}{\sigma}\left(\bar{Z}-\mu\right)$. We have

$$X' = \frac{\frac{\sqrt{n}}{\sigma}\left(\bar{Z}-\mu\right)}{\sqrt{\frac{W}{n-1}}}$$

and we now focus on the numerator. We have

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i.$$

Since the (Z_i) are independent $N(\mu, \sigma^2)$, we have that $\sum_{i=1}^n Z_i$ has a $N(n\mu, n\sigma^2)$ distribution, and hence $\bar{Z}$ has a $N(\mu, \frac{\sigma^2}{n})$ distribution. It follows that $Z' = \frac{\sqrt{n}}{\sigma}\left(\bar{Z}-\mu\right)$ has a $N(0,1)$ distribution. Hence,

$$X' = \frac{Z'}{\sqrt{\frac{W}{n-1}}}.$$

By applying **5.10**, using that $Z' \sim N(0,1)$ and $W \sim \chi^2_{n-1}$, we have that the distribution of X' is the t distribution with $n-1$ degrees of freedom.

Note that each Z_i has a $N(\mu, \sigma^2)$ distribution, so we would expect a statistic of the Z_i, such as X', to depend on μ and σ^2. In this case, however, we have shown that X' has a χ^2_{n-1} distribution, regardless of the values of μ and σ^2. This allows us to design statistical tests, using the statistic X', without needing to know (or estimate) μ or σ. One such test, is Student's t test.

7 Likelihood and Maximum Likelihood

Warm-up Questions

7.1 Let $f : \mathbb{R} \to \mathbb{R}$ by $f(\theta) = e^{-\theta^2+4\theta}$.

(a) Find the first derivative of f, and hence identify its turning point(s).

(b) Calculate the value of the second derivative of f at these turning point(s). Hence, deduce if the turning point(s) are local maxima or local minima.

Solution. We have $f'(\theta) = (-2\theta + 4)e^{-\theta^2+4\theta}$, which is zero if and only if $\theta = 2$. Further,

$$f''(\theta) = -2e^{-\theta^2+4\theta} + (4 - 2\theta)^2 e^{-\theta^2-4\theta}$$

so as $f''(2) = -2e^4 < 0$, which means that the turning point at $\theta = 2$ is a local maxima.

7.2 Let $(a_i)_{i=1}^n$ be a sequence with $a_i \in (0, \infty)$. Show that $\log\left(\prod_{i=1}^n a_i\right) = \sum_{i=1}^n \log a_i$.

Solution. This follows from iterating the formula $\log(ab) = \log a + \log b$.

Ordinary Questions

7.3 A sample of 3 is obtained from a geometric distribution X with unknown parameter θ. That is, $\mathbb{P}[X = x] = \theta^x(1 - \theta)$ for $x \in \{0, 1, 2, \dots\}$.

(a) Given this sample, find the likelihood function $L(\theta; 3)$, and state its domain Θ.

(b) Find the maximum likelihood estimate of θ.

Solution.

(a) The parameter θ of the geometric distribution takes values in $\Theta = [0, 1]$. The probability function of the geometric distribution is $p(x; \theta) = (1-\theta)^x\theta$, defined for $x \in \mathbb{N}$. Therefore, its likelihood function, defined for $\theta \in \Theta$, is given by $L(\theta; s) = (1 - \theta)^x\theta$. For the data point $x = 3$, this gives

$$L(\theta; 3) = \theta^3(1 - \theta).$$

(b) Differentiating, we have

$$\begin{aligned} L'(\theta; 3) &= -\theta^3 + 3\theta^2(1 - \theta) \\ &= \theta^2(-\theta + 3(1 - \theta)) \\ &= \theta^2(3 - 4\theta). \end{aligned}$$

We now look for turning points. Solving $L'(\theta; 3) = 0$ gives that $\theta = 0$ or $\theta = \frac{3}{4}$. Next, we see if these turning points are local maxima or minima. Since

$$\begin{aligned} L''(\theta; 3) &= 2\theta(3 - 4\theta) + \theta^2(-4) \\ &= -12\theta^2 + 6\theta \\ &= 6\theta(1 - 2\theta) \end{aligned}$$

it is easily seen that $L''(0;3) = 0$ and $L''(\frac{3}{4};3) = \frac{-18}{8} < 0$.
So, we know that $\theta = \frac{3}{4}$ is local maximum, but unfortunately since the second derivative is zero we don't know if $\theta = 0$ is a maximum or a minimum. However, since $L(0;3) = 0$ and $L(\frac{3}{4};3) > 0$ it must be $\hat{\theta} = \frac{3}{4}$ that is the global maximum, and hence also the maximum likelihood estimator of θ.

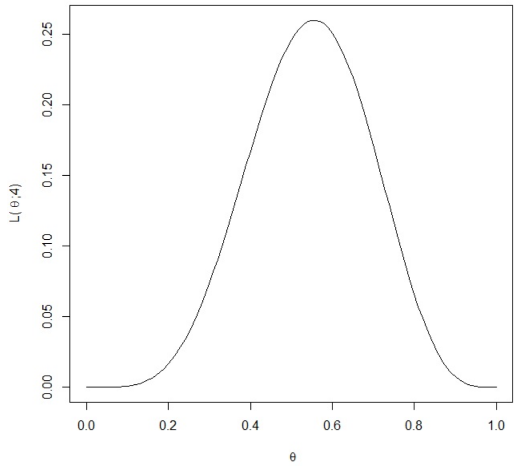
7.4 A sample of 4 is obtained from a $Bi(n, \theta)$ distribution.

(a) Assuming that n is known to be 9 and that θ is unknown, write down the likelihood function of θ, for $\theta \in [0, 1]$, given the (single) data point 4. Use a software package of your choice to plot a graph of it.

(b) Find the maximum likelihood estimate of θ, given this data point.

Solution. The probability function of the Binomial distribution is $p(x) = \binom{n}{x}(1-\theta)^{n-x}\theta^x$. The parameters n and θ take values in $\mathbb{N}$ and $[0, 1]$ respectively.

(a) Assuming $n = 9$, the likelihood function for θ is $L(\theta; x) = \binom{9}{x}(1-\theta)^{n-x}\theta^x$, defined for $\theta \in [0, 1]$. If x is observed to be 4, this gives $L(\theta; 4) = 126(1-\theta)^5\theta^4$. A plot in `R` gives



(b) From question **7.4**(a), the likelihood is $L(\theta; 4) = 126\theta^4(1-\theta)^5$. Differentiating,

$$\frac{dL}{d\theta} = 126\theta^3(1-\theta)^4(4(1-\theta) - 5\theta).$$

This will be zero when $\theta = 0$, $\theta = 1$ or when $4(1-\theta) = 5\theta$, i.e. when $\theta = \frac{4}{9}$. At this point, we could can look at the picture and note that $\frac{4}{9}$ is ' the turning point in the middle' and is therefore the global maximum.
Alternatively, we could try the second derivative:

$$\frac{d^2L}{d\theta^2} = 126(12\theta^2(1-\theta)^5 - 40\theta^3(1-\theta)^4 + 20\theta^4(1-\theta)^3).$$

At $\theta = \frac{4}{9}$ this is negative, so $\theta = \frac{4}{9}$ is a local maximum; at the other two it is zero, which is inconclusive, so we have to use the picture (or, numerically evaluate $L(\theta; 4)$ at the turning points) anyway - and the values we can see for of the likelihoods confirm that $\theta = \frac{4}{9}$ is the maximum.

7.5 A sample (x_1, x_2, x_3) of three observations from a Poisson distribution with parameter λ, where λ is known to be in $\Lambda = \{1, 2, 3\}$, gives the values $x_1 = 4, x_2 = 0, x_3 = 3$. Find the likelihood of each of the possible values of λ, and hence find the maximum likelihood estimate.

Solution. The probability of observing the values 4, 0 and 3 as independent observations from a $Po(\lambda)$ population is

$$\frac{\lambda^4 e^{-\lambda}}{4!}\frac{e^{-\lambda}}{0!}\frac{\lambda^3 e^{-\lambda}}{3!} = \frac{\lambda^7 e^{-3\lambda}}{144},$$

which gives a likelihood of 3.46×10^{-4} for $\lambda = 1$, 2.20×10^{-3} for $\lambda = 2$, and 1.87×10^{-3} for $\lambda = 3$. So the maximum likelihood estimate is $\lambda = 2$.

7.6 Given the set of i.i.d. samples $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, for some $n > 0$, write down the likelihood function $L(\theta; \mathbf{x})$, in each of the following cases. In each case you should give both the function, in simplified form where possible, and the parameter set Θ.

The data are i.i.d. samples from:

(a) the exponential distribution $Exp(\lambda)$ with $\lambda = 1/\theta$.

(b) the binomial $Bi(m, \theta)$ distribution, where m is known.

(c) the normal $N(\mu, \theta)$, where μ is known.

(d) the gamma distribution $Ga(\theta, 4)$.

(e) the beta distribution $Be(\theta, \theta)$.

Solution. The likelihood function is $L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta)$.

(a) The contribution to the likelihood from observation i will be $f(x_i; \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta} x_i}$. So the likelihood is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{1}{\theta} x_i} = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}.$$

The parameter set $\Theta = (0, \infty)$.

(b) The contribution to the likelihood from observation i will be $f(x_i; \theta) = \binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i}$, for $x_i = 0, 1, \ldots, m$. So the likelihood is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i} = \left[\prod_{i=1}^n \binom{m}{x_i}\right]\theta^{\sum_{i=1}^n x_i}(1-\theta)^{nm-\sum_{i=1}^n x_i}.$$

The parameter set $\Theta = [0, 1]$.

(c) The contribution to the likelihood from observation i will be $f(x_i; \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i-\mu)^2}{2\theta}}$. So the likelihood is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i-\mu)^2}{2\theta}} = \frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2}.$$

The parameter set $\Theta = (0, \infty)$.

(d) The contribution to the likelihood from observation i will be $f(x_i;\theta) = \frac{4^\theta}{\Gamma(\theta)} x_i^{\theta-1} e^{-4x_i}$. So the likelihood is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \frac{4^\theta}{\Gamma(\theta)} x_i^{\theta-1} e^{-4x_i} = \frac{4^{n\theta}}{\{\Gamma(\theta)\}^n} \left(\prod_{i=1}^n x_i\right)^{\theta-1} e^{-4\sum_{i=1}^n x_i}.$$

The parameter set $\Theta = (0,\infty)$.

(e) The contribution to the likelihood from observation i will be $f(x_i;\theta) = \frac{1}{B(\theta,\theta)} x_i^{\theta-1}(1-x_i)^{\theta-1}$. So the likelihood is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^n \frac{1}{B(\theta,\theta)} x_i^{\theta-1}(1-x_i)^{\theta-1} = \frac{1}{[B(\theta,\theta)]^n} \left(\prod_{i=1}^n x_i\right)^{\theta-1} \left[\prod_{i=1}^n (1-x_i)\right]^{\theta-1}.$$

The parameter set $\Theta = (0,\infty)$.

7.7 For each of **7.6**(a)-(e), find the log likelihood, and simplify it as much as you can.

Solution.

(a) $\ell(\theta;\mathbf{x}) = \log\left(\frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^n x_i}\right) = -n\log\theta - \frac{1}{\theta}\sum_{i=1}^n x_i$.

(b) We have

$$\begin{aligned}\ell(\theta;\mathbf{x}) &= \log\left(\left[\prod_{i=1}^n \binom{m}{x_i}\right] \theta^{\sum_{i=1}^n x_i}(1-\theta)^{nm-\sum_{i=1}^n x_i}\right)\\ &= \sum_{i=1}^n \log\binom{m}{x_i} + \sum_{i=1}^n x_i\log\theta + (nm - \sum_{i=1}^n x_i)\log(1-\theta).\end{aligned}$$

(c) $l(\theta;\mathbf{x}) = \log\left(\frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2}\right) = -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2$.

(d) We have

$$\begin{aligned}\ell(\theta;\mathbf{x}) &= \log\left(\frac{4^{n\theta}}{\{\Gamma(\theta)\}^n}\left(\prod_{i=1}^n x_i\right)^{\theta-1} e^{-4\sum_{i=1}^n x_i}\right)\\ &= n\theta\log 4 - n\log\Gamma(\theta) + (\theta-1)\sum_{i=1}^n \log x_i - 4\sum_{i=1}^n x_i.\end{aligned}$$

(e) We have

$$\begin{aligned}\ell(\theta;\mathbf{x}) &= \log\left(\frac{1}{[B(\theta,\theta)]^n}\left(\prod_{i=1}^n x_i\right)^{\theta-1}\left[\prod_{i=1}^n (1-x_i)\right]^{\theta-1}\right)\\ &= -n\log B(\theta,\theta) + (\theta-1)\left(\sum_{i=1}^n \log x_i + \sum_{i=1}^n \log(1-x_i)\right)\end{aligned}$$

7.8 The $IGa(\alpha, \beta)$ distribution is the distribution of $1/U$ if $U \sim Ga(\alpha, \beta)$.

Repeat **7.6** in the case where the data are i.i.d. samples from the inverted gamma distribution $IGa(1, \theta)$. Find and simplify the corresponding log likelihood.

Solution. If $U \sim Ga(\alpha, \beta)$ then we want the p.d.f. of $X = g(U) = 1/U$. The range of U is $\mathbb{R}^+$ so g can be considered as a decreasing function, with inverse $g^{-1}(x) = 1/x$, and $f_U(u) = \frac{\beta^\alpha u^{\alpha-1} e^{-\beta u}}{\Gamma(\alpha)}$ for $u > 0$, so we have

$$f_X(x) = f_U(g^{-1}(x)) \left| \frac{dg^{-1}(x)}{dx} \right| = \frac{\beta^\alpha x^{1-\alpha} e^{-\beta/x}}{\Gamma(\alpha)} \left| -\frac{1}{x^2} \right| = \frac{\beta^\alpha x^{-(1+\alpha)} e^{-\beta/x}}{\Gamma(\alpha)},$$

for $x > 0$ (and zero otherwise).

Hence, with $\alpha = 1$ and $\beta = \theta$, the contribution of observation i to the likelihood is $f(x_i|\theta) = \theta x_i^{-2} e^{-\theta/x_i}$, for $x_i > 0$. So the likelihood is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \theta x_i^{-2} e^{-\theta/x_i} = \theta^n \left(\prod_{i=1}^n x_i \right)^{-2} e^{-\theta \sum_{i=1}^n x_i^{-1}}.$$

The parameter set for θ is $\Theta = (0, \infty)$. The corresponding log likelihood is

$$\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}) = n \log \theta - 2 \sum_{i=1}^n \log x_i - \theta \sum_{i=1}^n \frac{1}{x_i}.$$

7.9 A set of i.i.d. samples $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is taken from a geometric distribution with unknown parameter θ, so that the probability function of X_i is

$$p(x_i) = \mathbb{P}[X_i = x_i] = (1 - \theta)^{x_i} \theta, \quad x_i = 0, 1, 2, \ldots, \quad 0 < \theta < 1.$$

Find the maximum likelihood estimate of θ based on the above sample.

Solution. The log-likelihood function is

$$\ell(\theta; \mathbf{x}) = \log \prod_{i=1}^n p(x_i) = \log(1 - \theta) \sum_{i=1}^n x_i + n \log \theta$$

The first derivative of ℓ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = -\frac{\sum_{i=1}^n x_i}{1 - \theta} + \frac{n}{\theta},$$

which is zero if and only if

$$\theta = \frac{n}{n + \sum_{i=1}^n x_i} \tag{7.1}$$

The second derivative of ℓ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{\sum_{i=1}^n x_i}{(1 - \theta)^2} - \frac{n}{\theta^2} < 0,$$

for all θ (because $x_i > 0$). So, $\hat{\theta}$ given by (7.1) is the required maximum likelihood estimate of θ.

7.10 Find the maximum likelihood estimate of θ when the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are i.i.d. samples from the binomial $Bi(m, \theta)$ distribution, where m is known, as in question **7.6**/**7.7**(b).

Solution. From **7.6**(b) we have

$$L(\theta; \mathbf{x} = \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] \theta^{\sum_{i=1}^n x_i}(1-\theta)^{nm-\sum_{i=1}^n x_i}.$$

and, as in **7.7**(b), this means that the log likelihood function is

$$\ell(\theta; \mathbf{x}) = \sum_{i=1}^{n} \log \binom{m}{x_i} + \sum_{i=1}^{n} x_i \log\theta + \left(nm - \sum_{i=1}^{n} x_i\right)\log(1-\theta)$$

The first derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = \frac{1}{\theta}\sum_{i=1}^{n} x_i - \frac{1}{1-\theta}\left(nm - \sum_{i=1}^{n} x_i\right)$$

and setting this equal to zero we get $\frac{1}{\theta}\sum_{i=1}^n x_i = \frac{1}{1-\theta}\left(nm - \sum_{i=1}^n x_i\right)$, which implies that $(1-\theta)\sum_{i=1}^n x_i = \theta nm - \theta\sum_{i=1}^n x_i$, and hence

$$\theta = \frac{1}{nm}\sum_{i=1}^{n} x_i$$

is the unique turning point of ℓ. The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{1}{(1-\theta)^2}\left(nm - \sum_{i=1}^{n} x_i\right)$$

Evaluating at our turning point, we have

$$\left.\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2}\right|_{\theta=\frac{1}{nm}\sum\limits_{i=1}^{n} x_i} = -\frac{1}{\theta^2}\sum_{i=1}^{n} x_i < 0.$$

Hence, this turning point is a local maxima. Since it is the only turning point, it is a global maxima, hence $\hat{\theta} = \frac{1}{nm}\sum_{i=1}^n x_i$ is the maximum likelihood estimate of θ.

7.11 Find the maximum likelihood estimate of θ when the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are i.i.d. samples from the normal $N(\mu, \theta)$, where μ is known, as in question **7.6**/**7.7**(c). Show that this estimator is unbiased (i.e. that $\mathbb{E}[\hat{\theta}] = \sigma^2$).

Solution. From question **7.6**(c), the likelihood function is

$$L(\theta; \mathbf{x}) = \frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2}$$

and so, as in **7.7**(c) the log-likelihood function is

$$\begin{aligned}\ell(\theta; \mathbf{x}) &= -\frac{n}{2}\log(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2 \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2.\end{aligned}$$

The first derivative of $\ell(\theta;\mathbf{x})$ is

$$\frac{d\ell(\theta;\mathbf{x})}{d\theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

and setting this equal to zero we have $\frac{n}{2\theta} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\theta^2}$, which implies that the unique turning point is

$$\theta = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2. \tag{7.2}$$

The second derivative of $\ell(\theta;\mathbf{x})$ is

$$\frac{d^2\ell(\theta;\mathbf{x})}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{1}{\theta^2}\left(\frac{n}{2} - \frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

and evaluating it at the turning point (7.2) gives

$$\left.\frac{d^2\ell(\theta;\mathbf{x})}{d\theta^2}\right|_{(7.2)} = \frac{1}{\hat\theta^2}\left(\frac{n}{2} - \frac{n}{\sum_{i=1}^{n}(x_i-\mu)^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) = -\frac{n}{2\hat\theta^2} < 0.$$

Hence (7.2) is the required maximum likelihood estimate $\hat\theta$ of θ.

For the last part, $\hat\theta$ is unbiased because, if the true value of θ is σ^2, then

$$\mathbb{E}[\hat\theta] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2.$$

Note that this is the usual estimator of the variance of a normal population with known mean. (The factor $\frac{1}{n}$ is correct; it applies in case where the mean is known, whereas the more commonly used $\frac{1}{n-1}$ applies to the case the mean is *unknown*.)

7.12 Find the maximum likelihood estimate of $\theta \in (0,\infty)$ when the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are i.i.d. samples from the gamma distribution $Ga(3,\theta)$.

If $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 3$, calculate the maximum likelihood estimate $\hat\theta$ and show that it does not depend on the sample size n.

Solution. The likelihood function is

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n}\frac{\theta^3}{\Gamma(3)}x_i^{3-1}e^{-\theta x_i} = \frac{\theta^{3n}}{2^n}\left(\prod_{i=1}^{n}x_i\right)^2 e^{-\theta\sum_{i=1}^{n}x_i},$$

defined for $\theta \in (0,\infty)$, and so the log-likelihood function is

$$\ell(\theta;\mathbf{x}) = 3n\log\theta - n\log 2 + 2\sum_{i=1}^{n}\log x_i - \theta\sum_{i=1}^{n}x_i.$$

The first derivative of $\ell(\theta;\mathbf{x})$ is

$$\frac{d\ell(\theta;\mathbf{x})}{d\theta} = \frac{3n}{\theta} - \sum_{i=1}^{n}x_i$$

and setting this equal to zero we get $\frac{3n}{\theta} = \sum_{i=1}^{n} x_i$, which means that the unique turning point is

$$\theta = \frac{3n}{\sum_{i=1}^{n} x_i} = \hat{\theta}. \tag{7.3}$$

The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{3n}{\theta^2},$$

which is negative for all θ. Hence, $\hat{\theta}$, defined by (7.3), is the required maximum likelihood estimate.

From $\bar{x} = 3$, we get $\sum_{i=1}^{n} x_i = n\bar{x} = 3n$ and so $\hat{\theta} = 3n/(3n) = 1$.

7.13 Suppose that a set of i.i.d samples $\mathbf{x} = (x_1, \ldots, x_n)$ is taken from a negative binomial distribution, so that each X_i has probability function

$$p(x_i) = P(X_i = x_i) = \binom{x_i + r - 1}{r - 1} \theta^r (1 - \theta)^{x_i}$$

for some "success" probability θ satisfying $0 \leq \theta \leq 1$, where $x_i = 0, 1, 2, \ldots$ and r is the total number of "successes". If r is known, find the maximum likelihood estimate of θ.

Solution. The likelihood function is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} p(x_i) = \theta^{nr} \prod_{i=1}^{n} \binom{x_i + r - 1}{r - 1} (1 - \theta)^{x_i}.$$

The log-likelihood function is

$$\ell(\theta; \mathbf{x}) = (nr)\log\theta + \sum_{i=1}^{n} \log\binom{x_i + r - 1}{r - 1} + \log(1 - \theta)\sum_{i=1}^{n} x_i.$$

The first derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = \frac{nr}{\theta} - \frac{\sum_{i=1}^{n} x_i}{1 - \theta}$$

and setting this equal to zero we get $\frac{nr}{\theta} = \frac{\sum_{i=1}^{n} x_i}{1-\theta}$ which implies that

$$\theta = \frac{nr}{nr + \sum_{i=1}^{n} x_i}. \tag{7.4}$$

The second derivative of $\ell(\theta; \mathbf{x})$ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{nr}{\theta^2} - \frac{\sum_{i=1}^{n} x_i}{(1 - \theta)^2},$$

which is negative, for all θ. So $\hat{\theta}$, given by (7.4), is the maximum likelihood estimate of θ.

7.14 As in question **7.4**, an observation from a $Bi(n, \theta)$ gives the value $x = 4$.

(a) If θ is known to be $\frac{3}{4}$ but n is unknown, with possible range $n \in \{4, 5, 6, \dots, \}$, write down a formula for the likelihood function of n and calculate its values for $n = 4, 5, 6$ and 7.

(b) Find the maximum likelihood estimator of n.

Solution.

(a) Assuming $\theta = \frac{3}{4}$, the likelihood function for n is $L(n; x) = \binom{n}{x}(\frac{1}{4})^{n-x}(\frac{3}{4})^x$, defined for $n \in \mathbb{N}$. If x is observed to be 4, this gives $L(n; 4) = \binom{n}{4}(\frac{1}{4})^{n-4}(\frac{3}{4})^4 = \binom{n}{4}\frac{81}{4^n}$. The requested values are:

	$L(n;4)$
$n = 4$	$\frac{81}{256} \approx 0.316$
$n = 5$	$\frac{405}{1024} \approx 0.396$
$n = 6$	$\frac{1215}{4096} \approx 0.297$
$n = 7$	$\frac{2835}{16384} \approx 0.173$

(b) The highest value in this table is for $n = 5$. To confirm that $n = 5$ gives the highest value of the likelihood over all integers $n \geq 4$, we can check the ratio

$$\frac{L(n+1;4)}{L(n;4)} = \frac{n+1}{4(n-3)},$$

which is less than 1 if $n \geq 5$. Hence for $n \geq 5$ the likelihood is decreasing in n, so 5 is the maximum likelihood estimate of n.

7.15 Suppose we have data $\mathbf{x} = (x_1, \dots, x_n)$, which are i.i.d. samples from a $N(\mu, \sigma^2)$ distribution, where μ is unknown and σ^2 is known.

(a) Find the log-likelihood function of μ.

(b) Show that the maximum likelihood estimator of μ is $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

(c) Let $k \in (0, \infty)$. Find the k-likelihood region for μ.

Solution.

(a) If $X \sim N(\mu, \sigma^2)$ then $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/\sigma^2}$. Hence, (by the same calculation as in Example 39), the likelihood is

$$L(\mu; \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right),$$

where the range of possible values for μ is $\Theta = \mathbb{R}$. So, the log likelihood is

$$\ell(\mu; \mathbf{x}) = -\frac{n}{2}\left(\log(2\pi) + \log(\sigma^2)\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

(b) We have

$$\begin{aligned}\frac{d\ell}{d\theta} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-1)2(x_i-\mu)\\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\\ &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i - n\mu\right),\end{aligned}$$

and

$$\frac{d^2\ell}{d\theta^2} = \frac{-n}{\sigma^2}.$$

Solving for $\frac{d\ell}{d\theta}=0$, we get $\sum_{i=1}^{n}x_i - n\mu = 0$, and hence the only turning point of ℓ is $\mu = (\sum_{i=1}^{n}x_i)/n$. As the second derivative is negative everywhere, this turning point is the global maximum. So the maximum likelihood estimator of μ is

$$\hat{\mu} = \frac{\sum_{i=1}^{15}x_i}{15}$$

(which, in this case, is the sample mean $\bar{x}$).

(c) The difference between the log likelihood at the maximum and at μ is

$$\begin{aligned}\ell(\bar{x};\mathbf{x}) - \ell(\mu;\mathbf{x}) &= \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - \sum_{i=1}^{n}(x_i-\bar{x})^2\right)\\ &= \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\mu^2-\bar{x}^2-2x_i\mu+2x_i\bar{x}\right)\\ &= \frac{1}{2\sigma^2}(n\mu^2-n\bar{x}^2)-2\mu n\bar{x}+2n\bar{x}^2\\ &= \frac{n}{2\sigma^2}(\mu-\bar{x})^2.\end{aligned}$$

So if we want to give an interval estimate containing those values of μ for which the log likelihood is within k of its maximum, we get

$$\begin{aligned}\frac{n}{2\sigma^2}(\mu-\bar{x})^2 &\leq k\\ (\mu-\bar{x})^2 &\leq \frac{2\sigma^2}{n}k\\ |\mu-\bar{x}| &\leq \sqrt{2k}\frac{\sigma}{\sqrt{n}},\end{aligned}$$

giving a traditional confidence interval with endpoints $\bar{x}\pm\sqrt{2k}\frac{\sigma}{\sqrt{n}}$.

7.16 As in question **7.4**(a) an observation from a $Bi(n,\theta)$ distribution with $n=9$ gives the value $x=4$.

(a) Find the range of values for which the log likelihood is within 2 of its maximum value. [You may do this either by inspection of a plot, or by using a computer package to solve the inequality numerically.]

(b) A 'traditional' approximate 95% confidence interval here would be of the form $\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$, where $\hat{\theta} = x/n$. Compare your answer to (a) to what this would give.

(c) Repeat this analysis with $n = 90$ and $x = 40$, and comment on your results.

Solution.

(a) The log likelihood is $\log\binom{9}{4} + 4\log\theta + 5\log(1-\theta)$, and the MLE is 4/9. So we want to solve

$$4\log\theta + 5\log(1-\theta) \geq 4\log\frac{4}{9} + 5\log(1-\frac{4}{9}) - 2.$$

Solving this in Maple gives the range of values $[0.161, 0.757]$.

(b) The approximate confidence interval calculation gives $[0.120, 0.769]$. This is similar, but a bit wider, especially at the lower end.

(c) The range of values with likelihoods within 2 of the maximum is now $[0.342, 0.550]$. The approximate confidence interval gives $[0.342, 0.547]$, which is very similar (slightly narrower at the top end).

7.17 A set of i.i.d. samples $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is taken from an inverse Gaussian distribution (known also as the Wald distribution), with p.d.f.

$$f(x) = \sqrt{\frac{\theta}{2\pi x^3}} \exp\left(-\frac{\theta(x-\mu)^2}{2\mu^2 x}\right) \qquad \text{when } x > 0,$$

and $f(x) = x$ for $x \leq 0$, with parameters $\mu, \theta > 0$.

(a) Assuming that μ is known, find the maximum likelihood estimate of θ based on the above sample.

(b) If both μ and θ are unknown, find the maximum likelihood estimate of (μ, θ) based on the sample given.

Solution.

(a) The log-likelihood function is

$$\ell(\theta; \mathbf{x}) = \frac{n}{2}\log\theta - \frac{1}{2}\sum_{i=1}^{n}\log(2\pi x_i^3) - \frac{\theta}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}$$

The first derivative of ℓ is

$$\frac{d\ell(\theta; \mathbf{x})}{d\theta} = \frac{n}{2\theta} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}$$

and for this to be zero we must have

$$\theta = \hat{\theta} = \frac{n\mu^2}{\sum_{i=1}^{n}(x_i-\mu)^2 x_i^{-1}}.$$

The second derivative of ℓ is

$$\frac{d^2\ell(\theta; \mathbf{x})}{d\theta^2} = -\frac{n}{2\theta^2} < 0,$$

for all θ and so $\hat{\theta}$ is the required maximum likelihood estimate of θ.

(b) The log-likelihood function is

$$\ell(\mu,\theta;\mathbf{x}) = \frac{n}{2}\log\theta - \frac{1}{2}\sum_{i=1}^{n}\log(2\pi x_i^3) - \frac{\theta}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}.$$

We now calculate the first partial derivatives. Differentiating with respect to μ and rearranging gives

$$\begin{aligned}\frac{\partial\ell(\mu,\theta;\mathbf{x})}{\partial\mu} &= \frac{\theta}{\mu^3}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i} + \frac{\theta}{\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)}{x_i}\\ &= \sum_{i=1}^{n}\frac{\theta}{\mu^3}\frac{x_i-\mu}{x_i}\left((x_i-\mu)+\mu\right),\end{aligned}$$

giving

$$\frac{\partial\ell(\mu,\theta;\mathbf{x})}{\partial\mu} = \frac{\theta}{\mu^3}\sum_{i=1}^{n}(x_i-\mu). \tag{7.5}$$

Differentiating with respect to θ gives

$$\frac{\partial\ell(\mu,\theta;\mathbf{x})}{\partial\theta} = \frac{n}{2\theta} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}. \tag{7.6}$$

For these to both be zero, from (7.5) we must have

$$\mu = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i,$$

and as before we must have

$$\theta = \frac{n\mu^2}{\sum_{i=1}^{n}(x_i-\mu)^2 x_i^{-1}},$$

so let

$$\hat{\theta} = \frac{n\hat{\mu}^2}{\sum_{i=1}^{n}(x_i-\hat{\mu})^2 x_i^{-1}}.$$

Then both first partial derivatives will be zero at $(\hat{\mu},\hat{\theta})$. The second partial derivatives are

$$\begin{aligned}\frac{\partial^2\ell(\mu,\theta;\mathbf{x})}{\partial\mu^2} &= -\frac{3\theta}{\mu^4}\sum_{i=1}^{n}(x_i-\mu) - n\frac{\theta}{\mu^3},\\ \frac{\partial^2\ell(\mu,\theta;\mathbf{x})}{\partial\mu\partial\theta} &= \frac{1}{\mu^3}\sum_{i=1}^{n}(x_i-\mu),\\ \frac{\partial^2\ell(\mu,\theta;\mathbf{x})}{\partial\theta^2} &= -\frac{n}{2\theta^2}.\end{aligned}$$

So the Hessian evaluated at $(\hat{\mu},\hat{\theta})$ is

$$\begin{pmatrix} -\frac{n\hat{\theta}}{\hat{\mu}^3} & 0\\ 0 & -\frac{n}{2\hat{\theta}^2}\end{pmatrix},$$

which is negative definite and so $(\hat{\mu},\hat{\theta})$ is a maximum. Hence $(\hat{\mu},\hat{\theta})$ is the MLE of (μ,θ).

Challenge Questions

7.18 The Pareto distribution has parameters $\alpha > 0$ and $\beta > 0$, and p.d.f.

$$f(x; \boldsymbol{\theta}) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \quad \text{when } x \geq \beta$$

and $f(x; \boldsymbol{\theta}) = 0$ when $x < \beta$. A set of n i.i.d. samples $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ are taken from a Pareto distribution, where the parameters $\boldsymbol{\theta} = (\alpha, \beta)$ are both unknown.

Find the maximum likelihood estimate of $\boldsymbol{\theta}$ based on the above sample.

Solution. The log-likelihood function is

$$\ell(\boldsymbol{\theta}; \mathbf{x}) = \log \prod_{i=1}^n f(x_i|\boldsymbol{\theta}) = -(\alpha+1)\sum_{i=1}^n \log x_i + n\log\alpha + n\alpha\log\beta,$$

assuming $\alpha > 0$, $\beta > 0$ and $\beta \leq x_i$ for all i. (If $\beta > x_i$ for any i then the likelihood is zero.) The first partial derivatives are

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \alpha} = -\sum_{i=1}^n \log x_i + \frac{n}{\alpha} + n\log\beta \tag{7.7}$$

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \beta} = n\alpha/\beta$$

(for $\alpha, \beta > 0$ and $\beta \leq x_i$ for all i).

Hence $\ell(\boldsymbol{\theta}; \mathbf{x})$ is increasing in β for as long as $\beta \leq x_i$ for all i, and otherwise equal to zero, so to maximise β we take $\hat{\beta} = \min(x_1, \ldots, x_n)$ (similar to Example 41). Then from (7.7) we have that

$$\frac{\partial \ell(\hat{\boldsymbol{\theta}}; \mathbf{x})}{\partial \hat{\alpha}} = 0$$

which implies that

$$\frac{n}{\alpha} = \sum_{i=1}^n \log x_i - n\log\hat{\beta}$$

and so there is a possible maximum at

$$\alpha = \frac{1}{\log\left(\prod_{i=1}^n x_i^{1/n} / \min(x_1, \ldots, x_n)\right)}.$$

Checking the second derivative,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \alpha^2} = -\frac{n}{\alpha^2} < 0,$$

for all $\alpha > 0$ and for all $\beta \leq x_i$.

So $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})$ is the required maximum likelihood estimate of $\boldsymbol{\theta}$.